

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:34:28 ; Search time 103 seconds  
(without alignments)  
5105.057 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVSPVLEVACKETQLI.....FISSVYAFDTLCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_25Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9514	100.0	1861	3	AAY90350 Drosophil
2	9482	99.7	1954	4	Abb62757 Drosophil
3	541	5.7	934	5	Abp64732 Human pro
4	408.5	4.3	726	4	Aab92933 Human pro
5	408.5	4.3	726	6	Abr48183 Human bla
6	319.5	3.4	526	4	Aab92906 Human pro
7	303.5	3.2	2779	4	Abb62371 Drosophil
8	300.5	3.2	1855	6	Abx47539 Breast ca
9	291	3.1	1742	7	ADB82767 Human pro
10	290.5	3.1	285	4	Aab92905 Human pro
11	284.5	3.0	1805	6	Abu35589 Protein e
12	284.5	3.0	1805	7	Abu36321 Protein e
13	277	2.9	1818	6	Abu36321 Protein e
14	277	2.9	1898	7	Aay30795 A human t
15	277	2.9	1898	7	ADD48869 Human pro
16	275.5	2.9	2688	4	Aam40883 Human pol
17	271.5	2.9	1975	4	Abb62094 Drosophil
18	270	2.8	2383	5	Abg65631 Human bre
19	269.5	2.8	2663	4	Aam39097 Human pol
20	269	2.8	1525	6	Aae32042 Human ves
21	268	2.8	2482	2	Aar72826 Human mit
22	268	2.8	2482	2	Aaw23996 Human mit
23	265.5	2.8	2633	4	Abg06505 Novel hum
24	264	2.8	2442	3	Aay77575 Human cyt
25	260	2.7	1979	7	Adb75595 Prostate

## ALIGNMENTS

RESULT 1	
AAAY90350	
ID	AAY90350 standard; protein; 1861 AA.
XX	XX
AC	AAAY90350;
XX	XX
DT	04-DEC-2000 (first entry)
XX	XX
DE	Drosophila Asp protein sequence.
XX	XX
Asp	Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor;
KW	tumour cell.
XX	XX
OS	Drosophila sp.
XX	XX
WO	WO200052478-A1.
XX	XX
08-SEP-2000.	
XX	XX
03-MAR-2000; 2000WO-GB0000785.	
XX	XX
04-MAR-1999; 99GB-00005007.	
XX	XX
(UYDU-) UNIV DUNDEE.	
XX	XX
Glover DM, Avides MDC;	
XX	XX
WPI; 2000-594203/56.	
DR	N-PSDB; AAA37761.
XX	XX
Use of Drosophila Asp polypeptide for identifying substances capable of	
disrupting microtubule organising center integrity and use of the	
identified substances for inhibiting mitosis in tumor cell.	
XX	XX
Claim 4; Page 43-44; 51pp; English.	
XX	XX
This sequence represents the Drosophila Asp protein. The invention	
relates to the use of Drosophila Asp polypeptide (or its homologue, or	
fragment) capable of stimulating formation and/or maintenance of	
microtubule organising centres (MTOCs), in an assay for identifying a	
substance capable of disrupting MTOC integrity. Asp polypeptide or its	
homolog is useful for identifying a substance capable of disrupting MTOC	
integrity. Substances identified by the method can be used to inhibit	
mitosis, e.g. in tumour cells	
XX	XX
Sequence 1861 AA;	
SQ	SQ

Query Match 100.0%; Score 9514; DB 3; Length 1861;

Db	1021	QKQAAA	SYIQMWRTYQLGRTORHEFLRQDRLLIMFVQREMRKSWMLQEKFEQOLKRA	1080
Qy	1081	AINIQOR	WRAKLSMRKCNADYLALRSSVLKQVAYPKATIQMRIDRNHYHSLRKNVICLOQ	1140
Db	1081	AINIQOR	WRAKLSMRKCNADYLALRSSVLKQVAYPKATIQMRIDRNHYHSLRKNVICLOQ	1140
Qy	1141	RLRAIMK	REORENLYRLRNASILVQKRYRMEQMIQDRNAVLRTRKCIINVQRRWRAATL	1200
Db	1141	RLRAIMK	REORENLYRLRNASILVQKRYRMEQMIQDRNAVLRTRKCIINVQRRWRAATL	1200
Qy	1201	QMERERK	NYLHLQTTTKRIQIKFRAKREMKQRAEFLOKKVTVLVQKRRALLQMKRER	1260
Db	1201	QMERERK	NYLHLQTTTKRIQIKFRAKREMKQRAEFLOKKVTVLVQKRRALLQMKRER	1260
Qy	1261	QEVYLH	REVITIKLQRRFHAQKSMRFRKAYRGTOAAVSCLOMHWNRHILRKRERNSFLQ	1320
Db	1261	QEVYLH	REVITIKLQRRFHAQKSMRFRKAYRGTOAAVSCLOMHWNRHILRKRERNSFLQ	1320
Qy	1321	ROAAITL	QRRYRARNLMIKOLKSYAQLOKAAITIQTRYRAKKAMQKQVLYQKQREAIK	1380
Db	1321	ROAAITL	QRRYRARNLMIKOLKSYAQLOKAAITIQTRYRAKKAMQKQVLYQKQREAIK	1380
Qy	1381	VQRRYR	GNLEMRKQIEVYQKQOAVIRLOKWRSTRDMRLCKAGYRRLRLSLSLTORKWR	1440
Db	1381	VQRRYR	GNLEMRKQIEVYQKQOAVIRLOKWRSTRDMRLCKAGYRRLRLSLSLTORKWR	1440
Qy	1441	ATVQARR	QOREIFLSTIRKVRMLQAFIRATLLMRQORREFEMKRAAVIQRFRACAML	1500
Db	1441	ATVQARR	QOREIFLSTIRKVRMLQAFIRATLLMRQORREFEMKRAAVIQRFRACAML	1500
Qy	1501	KARQDY	QLIOSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQOKFRGKRLMIEQRNCF	1560
Db	1501	KARQDY	QLIOSSVILVQKFRANRSMKQARQEFVQLRTIAVHLQOKFRGKRLMIEQRNCF	1560
Qy	1561	QLLRCS	MPGQARARGFMAKRFQOALMTPEMDLIROKRAAKVIORYWGVLIERRKHQ	1620
Db	1561	QLLRCS	MPGQARARGFMAKRFQOALMTPEMDLIROKRAAKVIORYWGVLIERRKHQ	1620
Qy	1621	GLLDIR	KRIARQAEKAVNSVRCKVQOAVRFLRGRTIASDALAVLSQDLRLSRTPVPHLL	1680
Db	1621	GLLDIR	KRIARQAEKAVNSVRCKVQOAVRFLRGRTIASDALAVLSQDLRLSRTPVPHLL	1680
Qy	1681	MWCSEF	MSTFCYGIMAAQAIRSEVDKQILIERCSRIILINLARYNSTVNTFQEGGLVTIAQM	1740
Db	1681	MWCSEF	MSTFCYGIMAAQAIRSEVDKQILIERCSRIILINLARYNSTVNTFQEGGLVTIAQM	1740
Qy	1741	LLRWCD	XDSIFNTLCTLIWVFAHCPKKRIIHDYMTNPEAIYMRVETKKLVAREKKMQ	1800
Db	1741	LLRWCD	XDSIFNTLCTLIWVFAHCPKKRIIHDYMTNPEAIYMRVETKKLVAREKKMQ	1800
Qy	1801	NARKPP	PMTSGRYKSOKINFTPCSLPSLEPDFGIIRYSPTTFISSVYAFDTILCKLOIDM	1860
Db	1801	NARKPP	PMTSGRYKSOKINFTPCSLPSLEPDFGIIRYSPTTFISSVYAFDTILCKLOIDM	1860
Qy	1861	F	1861	
Db	1861	F	1861	
RESULT 2				
ABB62757				
ID	ABB62757	standard; protein; 1954	AA.	
XX	AC	ABB62757;		
XX	AC			
DT	26-MAR-2002	(first entry)		
DE		Drosophila melanogaster polypeptide SEQ ID NO 15063.		
XX		Drosophila; developmental biology; cell signalling; insecticide;		
KW		pharmaceutical.		
XX		Drosophila melanogaster.		
OS				

XX WO200171042-A2.  
 XX 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06860.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 XX Disclosure; SEQ ID NO 15063; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS5773-  
 CC ABBS72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 1954 AA;  
 SQ  
 Query Match 99.7%; Score 9482; DB 4; Length 1954;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1855; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MELVWSPVLEVACKETLQIDNDRNFRKEVMIILKSKSNQPVKNRPKPTVVGKTLQKSGPT 60  
 DB 94 MELVWSPVLEVACKETLQIDNDRNFRKEVMIILKSKSNQPVKNRPKPTVVGKTLQKSGPT 153  
 QY 61 GAGTKMSVSAVQCKRMSAAAPPKQWVETAPSRPAWAHPPPOAPLVEKNVYKT 120  
 DB 154 GAGTKMSVSAVQCKRMSAAAPPKQWVETAPSRPAWAHPPPOAPLVEKNVYKT 213  
 QY 121 PQEEPVIYISQPSRLKENLSPMTPGNLLDVINDLRFPTLTETRGKQATIFPDNLAAPWT 180  
 DB 214 PQEEPVIYISQPSRLKENLSPMTPGNLLDVINDLRFPTLTETRGKQATIFPDNLAAPWT 273  
 QY 181 PTLKGNVKSANDMPRITPDDEDPATNKTFDVKGSETNINISLDTLDCSRIDGQPH 240  
 DB 274 PTLKGNVKSANDMPRITPDDEDPATNKTFDVKGSETNINISLDTLDCSRIDGQPH 333  
 QY 241 PLNKTITIVHATHALACIHEEGSPSPRTKSAIHDLDKDIKLVGSPLRKYSKMD 300  
 DB 334 PLNKTITIVHATHALACIHEEGSPSPRTKSAIHDLDKDIKLVGSPLRKYSKMD 393  
 QY 301 LSLSPQTKYATQSGMPNLNEMKIRSEIQNYRYYQEQIQIKAKDLNSSSSSEASLAGQOE 360  
 DB 394 LSLSPQTKYATQSGMPNLNEMKIRSEIQNYRYYQEQIQIKAKDLNSSSSSEASLAGQOE 453  
 QY 361 FLFNHSEILAQSSRNHLHVGRKSVKGSVPKNPHKRRSHLSFSDAPSNEIYRNETVAI 420  
 DB 454 FLFNHSEILAQSSRNHLHVGRKSVKGSVPKNPHKRRSHLSFSDAPSNEIYRNETVAI 513  
 QY 421 SPKQKQVEDTTLPRSAAPANASASSSAHAPHAQSKFKLAOTMSLMKKATPRKVRD 480  
 DB 514 SPKQKQVEDTTLPRSAAPANASASSSAHAPHAQSKFKLAOTMSLMKKATPRKVRD 573  
 QY 481 TSIQPSVKLYDSELYMQTCINPDPAATTTIDPFLASTWYLDQAVDRHQADFKKWLNAL 540

Db 574 TSIQPSVKLYDSELYMQTCINPDPAATTTIDPFLASTWYLDQAVDRHQADFKKWLNAL 633  
 QY 541 VSI PADLADANNKIDVGKLFNEVRNKELVAVPTKEBOSMNYLTKYRLETILKAAVELFF 600  
 Db 634 VSI PADLADANNKIDVGKLFNEVRNKELVAVPTKEBOSMNYLTKYRLETILKAAVELFF 693  
 QY 601 SEOMELPCSKVAVYVKNQALIRSDRNHLHDVVMORTILELLCENPLWRLGLEWVGE 660  
 Db 694 SEOMELPCSKVAVYVKNQALIRSDRNHLHDVVMORTILELLCENPLWRLGLEWVGE 753  
 QY 661 KIOMQSNREDIVGLSTFIINRLFRNKCEQRYSKAYTLTEEVAETIKKHSLOKILFLLPFL 720  
 Db 754 KIOMQSNREDIVGLSTFIINRLFRNKCEQRYSKAYTLTEEVAETIKKHSLOKILFLLPFL 813  
 QY 721 DOAKOKRIVKGNPCLFVKKSPHKETKIDILRPFSSSELLANIGDI TRELRRLGVLOHROTFF 780  
 Db 814 DOAKOKRIVKGNPCLFVKKSPHKETKIDILRPFSSSELLANIGDI TRELRRLGVLOHROTFF 873  
 QY 781 LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTQRLRVPAISRLOIRFNVKLAGALG 840  
 Db 874 LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTQRLRVPAISRLOIRFNVKLAGALG 933  
 QY 841 EANFOLGGDIAAQDIDVGHREKTLSELLMQLIYKFRSPKFAAATVLOKWRERHFWLVQ 900  
 Db 934 EANFOLGGDIAAQDIDVGHREKTLSELLMQLIYKFRSPKFAAATVLOKWRERHFWLVQ 993  
 QY 901 RRIIRKELMRHRAATVIOAVFRGHOMKYYKVLKFTERTQAAIILQKTRRYLAOKLYQ 960  
 Db 994 RRIIRKELMRHRAATVIOAVFRGHOMKYYKVLKFTERTQAAIILQKTRRYLAOKLYQ 1053  
 QY 961 SVHSITTIQRMWRAOOLGRQHRQRFVLEAREAFIQRITWRRELFAKLLAAATARLQRS 1020  
 Db 1054 SVHSITTIQRMWRAOOLGRQHRQRFVLEAREAFIQRITWRRELFAKLLAAATARLQRS 1113  
 QY 1021 QKQAAAASYIQOMWETIYQIGRIORHEFLRQDRLIMFVQRMESKWSMLEQKFEQOLKKA 1080  
 Db 1114 QKQAAAASYIQOMWETIYQIGRIORHEFLRQDRLIMFVQRMESKWSMLEQKFEQOLKKA 1173  
 QY 1081 AINIOORWRKLSMRKCNADYIALRSSVLKVQAYKATIQMRIDRNHYSLRKNVICLOQ 1140  
 Db 1174 AINIOORWRKLSMRKCNADYIALRSSVLKVQAYKATIQMRIDRNHYSLRKNVICLOQ 1233  
 QY 1141 RLRLTKMKREORENVLRLNASILVQRYRWEQMIQDRNAVLRTRKCTINVOBRWATL 1200  
 Db 1234 RLRLTKMKREORENVLRLNASILVQRYRWEQMIQDRNAVLRTRKCTINVOBRWATL 1293  
 QY 1201 QMRERKNYLHLQTTTKRIQIKFRAKREMKQRAEFLOLKKVTLVQKERRALLQMKER 1260  
 Db 1294 QMRERKNYLHLQTTTKRIQIKFRAKREMKQRAEFLOLKKVTLVQKERRALLQMKER 1353  
 QY 1261 QBYLHLREVTIKLQRRFHAQKSMRPMRAKYRGTOAAVSCLOQHWNRHLLRKERSFLQ 1320  
 Db 1354 QBYLHLREVTIKLQRRFHAQKSMRPMRAKYRGTOAAVSCLOQHWNRHLLRKERSFLQ 1413  
 QY 1321 RQAATTLORRYEARLNMILKOLKSYAQLKQAAITTIOTRYAKKAMOKOVVLYQKQREAIK 1380  
 Db 1414 RQAATTLORRYEARLNMILKOLKSYAQLKQAAITTIOTRYAKKAMOKOVVLYQKQREAIK 1473  
 QY 1381 VORRYRGNLEMRKQIEVYQKQRAVIRLQKWRWSIRDMLCKAGYRRIRLSLSIQKWR 1440  
 Db 1474 VORRYRGNLEMRKQIEVYQKQRAVIRLQKWRWSIRDMLCKAGYRRIRLSLSIQKWR 1533  
 QY 1441 ATVOARRQREIFLSTIRKVLMOAFIRATILMRQORREFEMKRAAVIQRFRACML 1500  
 Db 1534 ATVOARRQREIFLSTIRKVLMOAFIRATILMRQORREFEMKRAAVIQRFRACML 1593  
 QY 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLQKFKGRKLMIQRCNF 1560  
 Db 1594 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLQKFKGRKLMIQRCNF 1653  
 QY 1561 QLLRCSMPFGQARAGFMARKRFQALMTPEWMDLLIRQKRAAKVIQRYWRGYLIRRRQKHQ 1620

Db 1654 QLLRCSMPGFQARGFARMKRFOALMTPMMDLIRQRAAKVIQRYWRYLIRRRQKHQ 1713  
 QY 1621 GLDIDKRIQAQLQAEAKAVNSVRCKVQAEVFLRGRIASDALAVLSQIDRLSRTVPILL 1680  
 Db 1714 GLDIDKRIQAQLQAEAKAVNSVRCKVQAEVFLRGRIASDALAVLSQIDRLSRTVPILL 1773  
 QY 1681 MWCSEFMSTFCYGMQAIRSEVDKQLIERCSRIILMLARYNSTVNTFQGGIVTIAQM 1740  
 Db 1774 MWCSEFMSTFCYGMQAIRSEVDKQLIERCSRIILMLARYNSTVNTFQGGIVTIAQM 1833  
 QY 1741 LLRWCKDSEIFNTLCTLIWVFAHCPKPKRIIHDYMTNPAIYVRETCKLVARKERMKQ 1800  
 Db 1834 LLRWCKDSEIFNTLCTLIWVFAHCPKPKRIIHDYMTNPAIYVRETCKLVARKERMKQ 1893  
 QY 1801 NARPPMTSGRYKSKXINTPCSLPSLEDFGIIRYSPYTFISSVYAFPTILCKLQIDM 1860  
 Db 1894 NARPPMTSGRYKSKXINTPCSLPSLEDFGIIRYSPYTFISSVYAFPTILCKLQIDM 1953  
 QY 1861 F 1861  
 Db 1954 F 1954

## RESULT 3

ABP64732

ID ABP64732 standard; protein; 934 AA.

AC ABP64732;

XX 25-FEB-2003 (first entry)

DT Human protein SEQ ID 392.

XX Human; expressed sequence tag; EST; haematopoietic disorder;

XX central nervous system disease; viral infection;

KW peripheral nervous system disease; non-healing wound; infectious disease;

KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;

KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;

KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;

KW cytostatic; haemostatic; virucide; antibacterial; fungicide;

KW immunostimulant; cerebroprotective.

XX Homo sapiens.

OS WO200259260-A2.

XX 01-AUG-2002.

XX 16-NOV-2001; 2001WO-US042950.

XX 17-NOV-2000; 2000US-00714936.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;

PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-590824/63.

XX N-PSDB; ABQ99318.

XX New isolated polynucleotide, useful in research, diagnostic or

PT therapeutic methods, e.g. preventing or treating disorders involving

PT aberrant protein expression or biological activity.

XX Claim 20; SEQ ID NO 392; 394pp; English.

XX The present invention relates to novel human coding sequences (ABQ99268-

CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in

CC therapeutic, diagnostic and research methods. The polynucleotides may be

CC used in the field of molecular biology as hybridisation probes, primers

CC for PCR, for chromosome and gene mapping, for the recombinant production

CC of protein, or in generation of anti-sense DNA or RNA. The

CC polynucleotides are useful in diagnostics as expressed sequence tags

CC (ESTs) for identifying expressed genes or for physical mapping of the  
 CC human genome. The proteins may be used as molecular weight markers, or as  
 CC nutritional sources or supplements. The proteins may be used to maintain  
 CC and expand cell population in a totipotent or pluripotent state  
 CC useful for re-engineering damaged or diseased tissues, transplantation,  
 CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
 CC polynucleotides and proteins are useful for preventing, treating or  
 CC ameliorating disorders involving aberrant protein expression or  
 CC biological activity, e.g. haematopoietic disorders, central/peripheral  
 CC nervous system diseases, mechanical and traumatic disorders, non-healing  
 CC wounds, immune deficiencies and disorders, infectious diseases caused by  
 CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
 CC reactions and conditions, coagulation disorders, or cancer. The  
 CC polynucleotide sequences of the invention were assembled from ESTs  
 CC isolated mainly by sequencing by hybridisation, and in some cases,  
 CC sequences obtained from one or more public databases. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 934 AA;

Query Match 5.7%; Score 541; DB 5; Length 934;  
 Best Local Similarity 22.4%; Pred. No. 9.5e-32;  
 Matches 232; Conservative 188; Mismatches 350; Indels 266; Gaps 38;

QY 982 RQRFVRLREAAIFLO---RIWRRLFAKLLAAETARLQBSQKQAAAYIQMWRE-TY 1037  
 Db 4 RQLLRKHKASIVIOSTYRMYOYCFYQKL-----QWATKIIEKYRANK 48  
 QY 1038 QLGRIOREHFLRQDLIMFVQREMRSKWSMLEQRFQQLKRAAINIQWRRAKLSMRKC 1097  
 Db 49 KKQKVFQHNELKCTCV-----QAGFQDMNKKIQIEQHQAAIIQKHCKA-FKIRK- 99  
 QY 1098 NADYLALRSSVLKQV-AYRKATIQMDRIDRNHYSLR-KNVICLQORLAIMKWRQRENY 1155  
 Db 100 --HYLHRAFPVSIQRRYKRLT-----AVRTQAVICISQSYRG-FKVRKDIQN- 144  
 QY 1156 LRLRNASILVQKRYRQOQMIQDNAYLRTKCIINQREWRATLQWRERKNVHLQTT 1215  
 Db 145 --MHRATLIQSYRMRKVD---YETKTAIVVIQNTYRIVRVYVKKKKFLAVQKS 198  
 QY 1216 TKRIQIKRAKREMKQRAEFQLKKVTLVQVRRRLQWRKREQYHLREVTIKLQ 1275  
 Db 199 VRTIQAAFRGMK--VRQKLNVSSEKMAIV---NQSLCCYSKTSKTYEAVQSGVMIOE 253  
 QY 1276 RFHAQKSMRFRAKYRGTOAAVSLQMHWRNHLIRKE-----RNSF 1317  
 Db 254 WYKASGLACSQEAEYHSQSRAAVTIQKAFCRMVTRKLETKCAALRTIOFFLQMAVYRRF 313  
 QY 1318 LQLRQAATLQRRYRRLNNIKQLKSAQLKQAAITQTRYAKKAMQKOVLYQKQEA 1377  
 Db 314 VQQRQAATLQHYFRT-----WQTRKQFLLYRKAAVLQNHYRFLSAHQORVILQIRSS 369  
 QY 1378 IIVQRRYRGNLEMRKQIEVYQKQRAVIRLQKWRSDMR-LCK-----AGYR-- 1426  
 Db 370 VLIQARSGFIQKRK---FQEIKNSTIKIQAMWRRYRAKYLCKVKAACKIQAWYRW 425  
 QY 1427 -----RILSSLSTQKWRATVQARRQREIFLSTI 1456  
 Db 426 RAHKYIAVLAKVKIITQGCFTYTKLERTFLNVASAIITQKRWAILPAKIAHEHFLMIK 485  
 QY 1457 RKVRLMQAFIRATLLMRQRRREFEMKRAAVIQRFRPARCAMLKARQDYQLIQSSVILV 1516  
 Db 486 R-----HRAACLQAHYRG-----YKGEQVSLRQKSAALII 516  
 QY 1517 QRKFRANRSMKQARQBFVQLRTIAVHLQOKFRG---KRLMIEORNCQQLLR----- 1564  
 Db 517 QKVIAREAGKHERIKYIEFKKSTVILQALVRGMLVRKRFLEQRAKIRLLHFTAAAYHL 576  
 QY 1565 -----CSMFGFOARAGFMARKEF-QALMTPEMD--- 1593  
 Db 577 NAVRIQRAYKLYLAVNKNKQNVNSVICTQRFWRAR-----LQEKRFIQKYSIKKIEHGG 632



CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence and the combination of  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX

Sequence 726 AA;  
 SQ

Query Match 4.3%; Score 408.5; DB 4; Length 726;  
 Best Local Similarity 20.9%; Pred. No. 7,9e-22;  
 Matches 202; Conservative 147; Mismatches 321; Indels 297; Gaps 32;

QY 1594 --LIRQKRAAKVQIYRWGILLRRKQH--QGLLDIR-----KRIQLRQE 1635  
 DB 633 ECLSORNRAASVQKAVRHFLRKKQKFTSGIIKIALWRGYSWRKNDCTKIKAIRLS 692  
 QY 1636 KAVANSRCKVQAEVFLRGRFIASDALAVLSQDLRSNTVPHLLMWCSEFMSTFCYGM 1695  
 DB 693 LQVNR---ETREENKLYKRTALALHYLLTYKHLSAILEALKHLEVTR--LSPCCENM 747  
 QY 1696 AQARSEVDKQLIERCSR-----IILMARNSTVTWTFQGGHVTIAQMLL 1742  
 DB 748 AQSGAISKIFVLIRSCNRSIPCMVEIRYAVQVLLNVSKYEKTSAYDVENCIDILLEL 807  
 QY 1743 -----RWCDKDSEIFNTLCTLIWFAHCPKRIIHDWTNPETALYMYRETKKLVA 1793  
 DB 808 QIYREKPGNKVADKGSIFTKCCILAILL---KTNRASDVRSKVDRIYSLYKLTA 864  
 QY 1794 RKEKMKONARKPPPTWTSGRYSQK-----INFTPCS-----LPSLEPDPFGIIRYSPYTF 1842  
 DB 865 HKHKM--NTERI-----LYQKKNSSISIPETPTVTRIVSRKLPDVLRRDNMBEI 916  
 QY 1843 ISSVAFPTILCKLOI 1858  
 DB 917 TNPLQAIQVMDTLGI 932

RESULT 4  
 AAB92933  
 ID AAB92933 standard; protein; 726 AA.  
 XX  
 AC AAB92933;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:11593.  
 XX  
 KW Human; primer; detection; diagnosis; gene therapy; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 11593; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence and the combination of  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX

Sequence 726 AA;  
 SQ

Query Match 4.3%; Score 408.5; DB 4; Length 726;  
 Best Local Similarity 20.9%; Pred. No. 7,9e-22;  
 Matches 202; Conservative 147; Mismatches 321; Indels 297; Gaps 32;

QY 936 TERTQAIILOKFTRRYLAQKLYQSYHSI-ITIQWRRAQOLGRHQRFFVELREAAIF 994  
 DB 11 SEKMAAIVNQSAIACCYSRKTQ-YEAVQSEGVMIQEWYKASGLACQBAEYHSQRAAVT 69  
 QY 995 LQRIWRRLFAKLLAAAEATLQRSOKQQAASVIO--MQWRTYQLGRIQRIHFLRQD 1052  
 DB 70 IQKAF-CRMVTRKL-----ETQCAALRIQFFLOWAVY-----RRRVQOQKR 110  
 QY 1053 LIMFVQREMRKSWMLEQKEFQOLKRAAINIQWRRAKLSMRKCNADYLALRSSVLKVO 1112  
 DB 111 AAITLQHYFRT-W---QTRKQFLLYKAAVVLQNYHRAFLSAKHQORVYLQIRSSVILQ 166  
 QY 1113 AYRKATIQWRIDRNHYSLRNVICLOQLRAIMQWRQRENYLRLNASILVQ---KRY 1169  
 DB 167 ARSKGFIQKR-----KFOEIKNKSTIKIAMWRY 195  
 QY 1170 RMRQMIQDRNAYLSTRKCIINVO---RRWRATLQWRERKNYLHLQTTTKIQLKFRK 1226  
 DB 196 RAKK-----YLCKVRAACKIQAWYRCWRA-----HKEYLAVLKAVKTIQGCFTYK 240  
 QY 1227 REMKKQRAEFLQKKVTLVQKRRALLQMREROEYLHLREVTIKLQRRFHAQKSMRPM 1286  
 DB 241 LE---RTEFLNVRASAIILQKWRFAILPAKIAHSHFLMIKR----- 278  
 QY 1287 RAKYRGTOAAVSCLOMHWNRHLLRKRERNFLQLRQAATLQRRYRABNLMIKLSYQAQ 1346  
 DB 279 -----HRAACLIQAHYRG-----YKGRQVSLRQ----- 301  
 QY 1347 LKQAAITITRYRAKAMQKQVLYKQKOREALIKVQRRYRGNLENRKOIEVYQKORQAVI 1406  
 DB 302 -KSAALIIQKTYRAEAGKHERIKYIEFKSTVILQALVRGWLVRKFRLE-----QRAKI 355  
 QY 1407 RLOKWRSGIRDMRLCKAGYRRIRLSSLSIQKWRATVQARKQREFTLSTIRKVLMOQAFI 1466  
 DB 356 RL-----LHFTAAAYHNL----- 369  
 QY 1467 RATLIMRQOREFEMKREAAVVIQRRFRACMLKARQDYQLIQSSVILQKFRANRSM 1526  
 DB 370 -----AVRIQKAYKLYLAVKGNKQV-----NSVICIQKWFARLQOE 406  
 QY 1527 KQARQEFVQLRTIAVHLOQKFRGKELMIEQRNCFLLRCSMPGQARARQWARKFOAL 1586  
 DB 407 KRFTQKYSI-----KKIEHGEQECISQRNRAASVIOKAVRHFLRKK----- 449  
 QY 1587 MTPMMDLIRQKAAKV--IQRYVRGYLIRRRQKHQGLDIRKTAQLRQAKAVNSVRC 1644  
 DB 450 -----QKFTSGIIGIKIQLMRGYSWRKNDCTKIKAIRLSQVNRREIRENKLYK 500  
 QY 1645 KVQEAVERLGRFRTASDALAVLSQDLRSRTVPHLLMWCSEFMSTFCYGINAQAIRSEVD 1704  
 DB 501 RTALALHYLLTYKHLISAILEALKHLEVTR-----LSPCCENNAQSGAISKI 548

QY 1705 KQIERCSR-----IILNARYNSTVNTFQEGGLVTIAQMLL----- 1742  
Db 549 FVLIRSCNRIPQMEVIRYAVQVLLNVSKYEKTTSAVIDVENCIDILLELLQIYREKPGN 608  
QY 1743 RWCDDKSEIENITLCTLIWVFAHCPKKEKI IHDYMTNPEAIYVMVRETCKLVARKEKMKQNA 1802  
Db 609 KVADKGSITTKCCLLAILL---KTNRASDVRSRSKVDRIYSLYKLTAKHKM--NT 663  
QY 1803 RKPPMTSGRYKSKQ-----INTPCS-----LPSLEPDFGIRYSPYTFISSVAFDT 1851  
Db 664 ERI-----LYKQKNSSISIPETPVPTRIVSRILKPDWLLRRDNMBEITNPLQAIQM 717  
QY 1852 ILCKIQI 1858  
Db 718 VMDTLGI 724

RESULT 5  
ABR48183  
ID ABR48183 standard; protein; 726 AA.  
XX ABR48183;  
XX 12-JUN-2003 (first entry)  
DE Human bladder cancer associated protein sequence SEQ ID NO:82.  
XX Human; bladder cancer; cytostatic; gene therapy; vaccine.  
XX Homo sapiens.  
OS Homo sapiens.  
PN WO200303906-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 03-JUL-2002; 2002WO-US021338.  
XX  
PR 03-JUL-2001; 2001US-0302814P.  
PR 03-AUG-2001; 2001US-0310099P.  
PR 08-NOV-2001; 2001US-0343705P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
XX (BOSB-) BOS BIOTECHNOLOGY INC.  
XX  
XX Mack DH, Aziz N;  
PI  
XX  
XX WPI; 2003-201532/19.  
DR N-PSDB; ACC50994.  
DR  
XX  
XX Detecting a bladder cancer-associated transcript in a cell from a  
PT patient, comprises contacting a biological sample from the patient with a  
PT bladder cancer-associated polynucleotide or antibody.  
XX  
XX  
XX Claim 10; Page 261; 307pp; English.  
XX  
XX The present invention describes a method for detecting a bladder cancer-  
CC associated transcript in a cell from a patient. The method comprises  
CC contacting a biological sample from the patient with a polynucleotide  
CC that selectively hybridises to a sequence that is 80 % identical to a  
CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
CC encode the human bladder cancer-associated proteins given in ABR48146 to  
CC ABR48242). Bladder cancer-associated sequences from the present invention  
CC have cytostatic activities, and can be used in antisense gene therapy and  
CC in vaccine production. The method can be used for detecting a bladder  
CC cancer-associated transcript in a cell from a patient. The method is  
CC useful in diagnosing or treating bladder cancer and in screening for  
CC compounds that modulate bladder cancer, such as hormones or antibodies.  
CC The nucleic acid molecules from the present invention may be used in  
CC various screening and diagnostic methods, and for gene therapy, vaccine  
CC and/or antisense/inhibition applications  
XX

SQ Sequence 726 AA;

Query Match 4.3%; Score 408.5; DB 6; Length 726;  
Best Local Similarity 20.9%; Pred. No. 7.9e-22;  
Matches 202; Conservative 147; Mismatches 321; Indels 297; Gaps 32;

QY 936 TERTQAAIILQKTRRYLAQQLYQSYHSI-ITIQWWRQAQQLGRQHRQFVLELRAAIF 994  
Db 11 SEKMAIYNQSLCCVRSKTQ-YEAVQSEGVMIQEWYKASGLACQAEAEVHSQSRRAVT 69  
QY 995 LQIRWRRRLPAKLLAAETARLQSQQAASIIQ--MOWRTYQIGRIQRHEFLQRD 1052  
Db 70 IQKAF-CRMVTRKL-----ETQCAALRIQFFLQMAVY-----RRRFVQOKR 110  
QY 1053 LIMFVQRRMRSKWSMLEQRFQOLKEAANIQWRPAKLSMRKNADYLALSSVLKVO 1112  
Db 111 AAITLQHYFT-W---QTRKQFLLYRKAADVVLQNHYRAFUSAKHORVYLQIRSSVIIQ 166  
QY 1113 AYRKATIQMRIDRNHYYSIRKNVICLQORLRAIMKWRQRENLYRLFNASILVQ---KRY 1169  
Db 167 ARSGFIQKR-----KFEIKNSTIKIQAMWRY 195  
QY 1170 RMRQMIQDRNAVLRTRKCIINVQ---RRWRATLQMRERKNYLHLQTTTKRQIKPRAK 1226  
Db 196 RAKK-----YLCVKAAACKIQAWYRCWRA-----HKEYLAVLKAVKIIQGCFTYK 240  
QY 1227 REMKQRAEFLQKVTLLVQKERRALLQMKERQEVYHLREVTIKLQRFHAKSKMRPM 1286  
Db 241 LE-----RTRFLNVRASAIILQKWRATILPAKIAHEHFLMTKR----- 278  
QY 1287 RAKYRGTAQAVSCLQMHWRNHLKRERNSFLOLQRAAITLQRRYRARLNMKILKLSYAQ 1346  
Db 279 -----HRAACLIQAHYRG-----YKGRQVSLRQ----- 301  
QY 1347 LKQAAITIQTRYPAKKAMQKVVLVQKREAILIKVQRRYRGNLEMRKQIEVYQKQRAVI 1406  
Db 302 -KSAALTIQYIRAREAGKHERIKYIEFKKSTVILQALVRGWLVRKFFLE-----QRAKI 355  
QY 1407 RLQKWRSIRDMLCKAGYRIRLSSLSIQKWRATVQARRQREIFLSTIRKVLMOAFI 1466  
Db 356 RL-----LHFTAAAYVHLN----- 369  
QY 1467 RATILMRQRRBEMKRRRAAVIQRFRPARCAMLKARQDYQLIQSSVILVQRKFRANRSM 1526  
Db 370 -----AVRIQRAVKLVAVKNANKQV-----NSVICIQRFARLQE 406  
QY 1527 QKARQEFVLRTTAVHLQKFRGKRLMIEQRNCPOLLRCNMPGFQAPARQFARKRFQAL 1586  
Db 407 KRPIQKYHSI-----KKEHEGQECLSQRNRAASVIOKAVRHFLRKK----- 449  
QY 1587 MTPFEMDLIRQKRAKV--IQRYVRGYLIERROKHQGLLDIRKIRLAQLRQEAQAVNSVRC 1644  
Db 450 -----QEKPTSGIILKIQALWRGYSWRKNDCTKIKAIRLSLQVNVREIRENKLYK 500  
QY 1645 KVOEAVRFLRGRFIASDALAVLSQLDRLSRTVPHLLMWCSEFMSTFCYIGIMAQIRSEVD 1704  
Db 501 RTALALHLLTYKHLISAILALAKHLEVTR-----LSPLCCENMAQSGAISKI 548  
QY 1705 KQIERCSR-----IILNARYNSTVNTFQEGGLVTIAQMLL----- 1742  
Db 549 FVLIRSCNRIPQMEVIRYAVQVLLNVSKYEKTTSAVIDVENCIDILLELLQIYREKPGN 608  
QY 1743 RWCDDKSEIENITLCTLIWVFAHCPKKEKI IHDYMTNPEAIYVMVRETCKLVARKEKMKQNA 1802  
Db 609 KVADKGSITTKCCLLAILL---KTNRASDVRSRSKVDRIYSLYKLTAKHKM--NT 663  
QY 1803 RKPPMTSGRYKSKQ-----INTPCS-----LPSLEPDFGIRYSPYTFISSVAFDT 1851  
Db 664 ERI-----LYKQKNSSISIPETPVPTRIVSRILKPDWLLRRDNMBEITNPLQAIQM 717  
QY 1852 ILCKIQI 1858  
Db 718 VMDTLGI 724

RESULT 6

AAAB92906  
ID AAB92906 standard; protein; 526 AA.

XX AC AAB92906;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:11533.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 11533; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention

XX SQ Sequence 526 AA;

Query Match 3.4%; Score 319.5; DB 4; Length 526;

Best Local Similarity 20.5%; Pred. No. 2.9e-15;

Matches 118; Conservative 92; Mismatches 124; Indels 243; Gaps 16;

XX 805 VEVILLRDDITRQLRVPAISRLQRIENFVNLGALGEANFQL---GGDIAAQDIVDGRH 860

Db 1 MELLTONWDLSSKLRIPAISSRLQRMHNVDIVLQVLKSGRIELSDHGNLTLSKDIVDRHR 60  
Qy 861 EKTLSLAWQLIYKFRSP-----KPHAAATVLQKWRR 892  
Db 61 EKTLLRLWKTAFQVDFISLNLQDLKEIAPLKHFKTSIKKTIISLSSCHSDLLINKKGR 120  
Qy 893 -----HWLHV----- 898  
Db 121 DSGSFEQYSENIKLMDWVNAVCAFYKNKVENFTVPSDGRVLCYLIHHYHPCVPLDAI 180  
Qy 899 ----- 898  
Db 181 CORTTQVTECTQGTGVSVLNSSSDSLDMSLKAFDHENTSELYKELLEKNKPHLVR 240  
Qy 899 -----IQRRIRHK-----ELMRHRAATVIQAVFRGHQ 926  
Db 241 SAVRDLGGIPAMINHSMSNTIPDEKVITVLSFLCARLLDLRKEIRARLIQTWRKYK 300  
Qy 927 MRKYVKLFKTERTOAAIILQKFTTRYLAOKLYOSYHSIITQWRWRAQQLGRQHRQFV 986  
Db 301 LKTDLKRHQ-EREKAARIQLAVINFLAKQRLRKVNAAALVIQYV----- 345  
Qy 987 ELREAAIFLQIRWRRLPAKLLAAAFARLQSQKQQAASYYIQWRTYQLGRQHE 1045  
Db 346 -----RRVLAQRKLLMLKKEKLEKQNK--AASLIQGYWRRYS---TQR 385  
Qy 1047 ELRQDLIMFYQRRMRSSKWSMLEQKFEQOLKRAAINIQORWRAKLSMRKNADYALRS 1106  
Db 386 FLKLIYSIILQSRIR-----MLIATSYKRVLWATVIQRHWRVLRKQQRQYEMLS 441  
Qy 1107 SVLKVQA-YRKATIQRIDRNHYSLRXNVICLQORLEA-IMKMRQRENVYLR-----N 1160  
Db 442 STLIIQSMEFK-----WKQRK---MQSQVKATVILQAFREWHLRKQAKSEN 485  
Qy 1161 ASILVQKRYMRQMIQDRNAVLRTRKCIINVRWR 1197  
Db 486 SALLIQSWRMHKL-----RKYYIRSCVVIQKFR 518  
RESULT 7  
AAB62371  
ID AAB62371 standard; protein; 2779 AA.  
XX AC AAB62371;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 13905.  
XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX PR 11-JUL-2000; 2000US-00614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX PI WPI; 2001-656860/75.  
XX DR N-PSDB; ABL06474.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.  
PT











ABU35589  
ID ABU35589 standard; protein; 1805 AA.  
XX AC ABU35589;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #21116.  
XX XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX XX Mycoplasma genitalium.  
XX XX WO200277183-A2.  
XX XX 03-OCT-2002.  
XX XX 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 28-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
XX N-PSDB; ACA39459.  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA39459.  
XX XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX XX Claim 25; SEQ ID NO 63513; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX XX Sequence 1805 AA;  
XX QY 296 ESMKDLSPQTKY-----AIQSMFNLEMKIRSTEONRYIQ-----EQOIQIK 341  
XX DB 208 ENOKRLISIEYENTYRELVSADNELQNVYENIDQNIQFKHQYQTYRDELSQLEKIKLT 267  
XX QY 342 AKDINSSSEASLACQOEFLEHSEILAOSSRFNHEVGKSVKSGPVKNPKRRRSHL 401  
XX DB 268 KQELVKEKSAURVKIDDADFIYINAR--LAE-----LDDVAK-----QL 303  
XX QY 402 SFSD--APSNESLYRNETVAISPPK---KQYEDTTLPSA-APANASARSSAHAWPH 454  
XX DB 304 SFQGITKQNAQHVDEKLVALNKEKDRLNTQKEAFNLRQSALIDINKLOQENELFA-KH 362  
XX QY 455 A--QSKFKFLAQTWSLMK-----KPAIPRKVYDTSIQSVKLYDSELYMOTCINPDPFAAT 508  
XX DB 363 LEHQQNEFEQKQSDSLKLETEYKALQHKINEFKNESATK--SEELNQ---ERELPEKR 417  
XX QY 509 TTIDPEL--ASTWYLDL-----QAVDRHQADFKKWLNALVSIADLADLNL-----NKIDV 557  
XX DB 418 REIDTLLTQASLEHYEHORESSQLLKDKQNEVKOHFQNLVYAKKELDKERNLLDQKKVDS 477  
XX QY 558 GLFNEVRNKELYVAPTRPEQSMNYLTXYKLETLRKAVALFELFSEQMR-----LPC 608  
XX DB 478 EALF---QLKEKVAQERKELEEL--YLVKQKQDKQENEL--LFFEKQLKQHQADPENELEA 532  
XX QY 609 SKVAVYVVKQALRIS-----DNLHLDVVMQRTILELLCFNPLWLRGLGVVFGKIQ 663  
XX DB 533 KQQLFEAKHALE-RSFIKLEDEKDLNTRAQO-----TANEFSSQ 571  
XX QY 664 MQSNRDIIVGLSTFILNRLFRN--KCEQRYSKAVTLTEVYAEITKKHSLQKILFLLPDLQ 722  
XX DB 572 LKTDKXSADFELMLQNEVENLOQEKQKLFQERTYFERNAAVLNRLNQQKREEL-----626  
XX QY 723 AKQRIIVKHNPCLFVKKSPHKTKDILLR-FSEELIAN-----IGDITRELRRLGYVL 774  
XX DB 627 -----QKQETLDQTKSFEQERLINQREHKELVASVEKQKELGKGL 668  
XX QY 775 Q-HRQTFLEDFDYAFNNLA---VDLRGVLTRVVEVILARDLDTLQRLVPAISRLQRI 830  
XX DB 669 QDFSQTSLN---ASKNLAEREMAIKFEKEIEATEKQLND-----VN 708  
XX QY 831 NVKLALGALCEANFOLGSD-----IAAQDIVDGHREKXTLSILLWOLYKFRSPKPHAAATV 885  
XX DB 709 NAEVQADIALQALQNSLNQERSELQNAKQRIADFENDSLKL-----749  
XX QY 886 LQKWRRHVLHVVIQRRIRHELM-----RRHRAATVIOAVFRGHQMRKYVLFKTERTOA 941  
XX DB 750 -----NEVELSLQKELQELQTLLEANQKQHSYQN--QAYFEG-----ELDKLNEKQ 793  
XX QY 942 AI--ILQKFTFRVLAQKOLYQSVHSIITIQRWRAQQLGHRQRQFVELREAAFLQRIW 999  
XX DB 794 AFLNLRKQKQMEVDAIKQRLSDKHQALNMQO-----ABLDKRTH-----ELNNA--FLNHDA 843  
XX QY 1000 RRLRFAKLLAAAEATAR---LQRS---QKQAAASYI---QMWQ-----RTYQLGRI 1042  
XX DB 844 DQKSLQDLATVTKETQKLDLERSALLEKQREFAENVAGFKRHSNKTSLQKIYELTKK 903  
XX QY 1043 QRHEFLQRDLIMFVQORMSKVSMLEQRK--BFQQLKRAINIQQWRKAKLSNRKNAD 1100  
XX DB 904 QESQTKETELKIAFSDQLQDYQVFELOKQDEFRQ-----IEAKQELDKLA-EKNQV 957  
XX QY 1101 YLALRSSVLVQAVRKATIQRID-RNHYSYLRKNVICLOQLRAIMKMBEQ-----1151  
XX DB 958 KLELDNRFQALQKQKQDTVQAQLELEREQHQLNLEQTAFQANESILKQRELTQKIQAF 1017  
XX QY 1152 -----RENYLRLENASILVQKRYRMQOMIQDRNAYLTRKCIINVOQRWRATLQWRE 1205  
XX DB 1018 HYELKKRNQFALAKGKELFAKQDQQRKD--QEIN-----WR-----FKQF 1056  
XX QY 1206 RKNYLHLQTTTKR-----IQIKPRAKEMKKQRAEFLQLKVVILVQKRRAL 1253





Db 387 -BTEHKQL-----QKRIEFGKIESAKSEALLIQERELL-----EKRREIDLLTQAS 433  
QY 438 APANASARSSAHAPH-----AOSKKFKLAQTMSLMKPA 473  
Db 434 LEYEOQRRTNOVLKQKHROVQHFQNLVHAKKLDQKRHYLABQKRIDEQIFKLEKETA 493  
QY 474 TPRKVRDTSIQSVKILDSYELMYQTICINPDPAATTIDPFLASTMYLDEQAVDRHQAD- 532  
Db 494 TERRELE-----KLY-----LVKQKQDQKENDL 517  
QY 533 ---FKKWLNALVSI PADLADLANNKIDVGLFNEVRNKLIVAPTKEEQSMNYLTKYRLET 590  
Db 518 LIFELEQLR-----QYQADFENEI-----EEKQNELFASQKSLQKFTQKLNKEAEL 563  
QY 591 LRKAAVELEFFSEQMRLPCSKVAVYVYKQALRTSRDNLHLDDVVMQRTILELLICFNPLWL 650  
Db 564 NOKA-----QKIA-----EDWAHLKQNKHHAD-----L 587  
QY 651 RLGLVVFGEKIQMOSNRDIVGLSTF-----ILNRLFRNKCEE-----OKYSKAYTL 697  
Db 588 EIFLEGEFNH-LQOEKHKLEARTQFDNRVSLLSARFKQKQALVKQKOSLEQLTAAFNK 646  
QY 698 TEYEAETIKKHSLOKILFLLPFLDQAKRIIVKHNPCLFVKKSPHKETK-----DILLRFS 753  
Db 647 EQEAVERDWKRLANL-----EKQEMLGDKVHQFDENSLNISKLAERELAIFK 697  
QY 754 SELLA-----NIGDITRELRLGYVLQ-----HROTFLDFEDYAFNLAVD 794  
Db 698 EKELBAAQKQLSDNNNAGLKLDKLSKTERLELEASKERILDFYDESSRIA-D 756  
QY 795 LRDGVRLLTRVVBVILLRDDLTQLRLVPAISRQIRFNVKLAIGALCEANFOLGCGDIAAD 854  
Db 757 YESDLQ-ARLAEVKTLEKN---QOETAASERE-----LKVALEKLNQA-----796  
QY 855 IVDGHRKTLTLLWQLIYKFRSPKFAHAAATVILQKWRHMLHVITQRRIRHKELMRHRA 914  
Db 797 -----KKAFLQIRKQQLLEIASVK-----QQLAQKANLLKNQQA 830  
QY 915 ATVIQAVFRGHQVRKVKVLFKTERTQAAIILQKFTRYLAQKQLQSYHSIITIQRWFA 974  
Db 831 -----BLDK-----QTELEAFAFLQDQTDK-----KELEKALHSV-----860  
QY 975 QQLGQHQRORFVELREAAIFLQIRWRRLFAKLLAAETARLQSQKQOAAAASYIQMOW 1034  
Db 861 -----KSKQELLE-RERSFLQK-----QREFAHVAGFRQVHF 894  
QY 1035 RYVQLGRI-----QRHEFLQRDL-IMPVQRRMKSWMLEQKQKPFQOLKRAAINIQ 1086  
Db 895 KTTOMORLSEFNKQSQSEQIKRETELKIAFADIKKDYQLFELQKQNEFQO-----IRQ 947  
QY 1087 RWEAKLSMRKCNADYIALRSSVLKQAYRKA-----TIQMRID-----RNHYISLRN 1134  
Db 948 KHK-----ELELLAQKQALQKQELQKATALASQDQDTVQAKLDLARQOHELELRQN 999  
QY 1135 VTCLOQLRAIMKMRQRENRLRNASLIVQKRYR---MRQMIODRNAYLRTRKCIIN 1191  
Db 1000 ---AFNQASISLNKQRLQTNQVYVHGL- KQGEKHLTKORLLAEKQKHQDAEIN 1055  
QY 1192 VQRRWRATLOMRERKNRYLHLOTTTKR-IQIKPRAKREMKQRA-----BFLQJ 1239  
Db 1056 -QR-----FKQFENEYADFQAKKRELQELNQIRRNLEQSNASLLKKRNQJLTLDL 1107  
QY 1240 KKVTLVQKRRALLQMRKERQBYHLRVTIKLQRRFHAQKSMRFPRAKYRGTO-----1294  
Db 1108 RKVQHTQTNR---VQINTQIKFLEKKNFKQASDEAALQKALLIKRLRSFASKLOLQ 1164  
QY 1295 ---AAVSCLOMWRNHLRKRENSFLQQAATTQLRRYARLNMTKOLKSYAQ-----1346  
Db 1165 EALAIQKLEFDEKDEQKSEINNAKQLQKQKLEKQNFQAKQKQLEFQDQCORLDVEK 1224  
QY 1347 ---LQQAATITQYRAKQAMQVLYQKQREAIIVQRRYRGNLBRKQIE-----1396  
Db 1225 RLLKQKVLQKLNLSKSVLYTKNRADLSQQO-----LQHKYANLLEKELQKQAKRALDK 1278

QY 1397 ---VYQORQAVIRLQKWRISIRDWRLCKAGYRIRLSSLSIOR-----KWRATVQARQ 1448  
Db 1279 KIRALIYKGAQVASEL-----ROKKQLLSAQKQVDDKSLLEQOQRH 1321  
QY 1449 REIFLSTIRKVELMOAFIRATILMRQORREFEFMKRRAAVVIQRRFRACMLKARQDYOL 1508  
Db 1322 LQNLSSETKKKQES-----LEHDINKFPQRKEAV-----1351  
QY 1509 IOSSVILVORKPRANRSMKQARQBFVQLETTIAVHLQOKFGRKRLMLEORNCFOLLRCSP 1568  
Db 1352 ---SSIL-----NSHKKLKQKQEGELQGI-----OKLSLKKQTQEQE--FSKL-----1389  
QY 1569 GFQARARGFMARKRQALMTPEWMDLIRQKRAAKVIQRYWRGYLIRRRKQKHGLDI---1625  
Db 1390 -YQOREK-----LDQRTTSLKHLRELKQANEATAHKNREVIENY 1430  
QY 1626 -RKRIAQLRQEAKAVNSRVCKVQEAVERFLRGRFIASDA--LAVLSQDLRLSRTVPHLLMW 1682  
Db 1431 YKKEIQLTTEKSEFDNNKNRLEFVFRKIRNIEKKEAHIKTVLETKKR---HL---1483  
QY 1683 CSEFMSTFCYIGIMAAQAIRSEVDKQI-----ERCSRIILNARYNSTVTNTFOE 1731  
Db 1484 -----VETEAVKLHLQKQSIISKQBELKEIKERSVRDISHNTNKQREBELSLHQ 1532  
QY 1732 GGLV--TIAQMLLRWCDKDXSEIFNTLCTLIWVFAHCPKKRIIH--DYMTNPEAIYMWRE 1787  
Db 1533 NKLLQKNUAERRETNKNDLSLTQKIQTAKQKLE--KEARILKLEKRAVEQYQQA-E 1589  
QY 1788 TKKLVARKEKMQNARKP--PPMTSGRYKSQKINFTPCSLPSLEPDF 1832  
Db 1590 ITRLXTRNADLEKNDKNHLPPLF-----KINGNDMNPYPYPWF 1629

## RESULT 14

AAV30795  
ID AAY30795 standard; protein; 1898 AA.  
XX AAY30795;  
AC AAY30795;  
XX 25-NOV-1999 (first entry)  
DT A human trichohyalin (TRHY) protein.  
DE Human trichohyalin (TRHY) protein.  
KW Human; trichohyalin; TRHY; protein; tissue structure; wound healing;  
KW terminally differentiating epidermal tissue; proteinaceous gel;  
XX breast implant.  
OS Homo sapiens.  
XX US5958752-A.  
PN 28-SEP-1999.  
PD 14-FEB-1997; 97US-00800644.  
PF 30-APR-1993; 93US-00056200.  
PR (USSH) US DEPT HEALTH & HUMAN SERVICES.  
XX Kim I, Chung S, Park S, Steinert PM, Lee S;  
PI WPI; 1999-561041/47.  
DR N-PSDB; AA22301.  
XX Human trichohyalin useful for forming a proteinaceous gel that promotes  
PT wound healing.  
XX Disclosure; Fig 3A-W; 126pp; English.  
PS The present sequence represents a human trichohyalin (TRHY) protein. The  
CC protein is found in terminally differentiating epidermal tissue, and is  
CC involved in forming the structural architecture of such tissue. The

CC trichohyalin protein is useful for forming a proteinaceous gel which may  
 XX then be used for healing wounds, or in breast implants  
 SQ Sequence 1898 AA;

Query Match 2.9%; Score 277; DB 2; Length 1898;  
 Best Local Similarity 21.7%; Pred. No. 3.2e-11;  
 Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

QY 891 RRHLVHVVIQRIIPKELMRHRAATVIOAFVGHOMKVKVLFKTFRTQAAILQKFTF 950  
 DB 648 RRHLLKSEQERHRLERHRAATVIOAFVGHOMKVKVLFKTFRTQAAILQKFTF 950  
 QY 951 RYLAQKOLYQSYHIT-IQKW-WR-----AQQLG----- 978  
 DB 702 QELAEEQEQARERIKSRIPKQWQLESEADARQSVLLLEAPQAGRAEAPEQEKERRE 761  
 QY 979 -----RQHRQRFVELREAAIFLQRIW-----RRRLPAKLLAAETARLORS 1020  
 DB 762 SELQWQEEERAHQQOEEBQRRDFTWQ--WQAEKSEKRGRLSARPLREQRERQRAE 819  
 QY 1021 QKQQAASVYIQMWRTYQIGRIQSH-----EFLRQDLIMFVQRRMSEKSMLEQ--RKE 1073  
 DB 820 ERQREORFLPEEKEKQGRQRREREKEHQLFLEEEQL---QRREAAQQLQEEEDGLQE 876  
 QY 1074 FQOLKRAAINIQWRRAKLSMRKCNADYLAIRSSVLKVQYRKATIQWRIDRNHYSLRK 1133  
 DB 877 DQERRRQEQRRDQKRWQLEERKRRH-----TLYAKPALQEQ-----LRX 918  
 QY 1134 NVICLOQLRAIMKWRQRENYLRLNASILVKRYRMQMIQDRNAYLTRKCIINVQ 1193  
 DB 919 EQQLLOEBEEL--QREERKRRQEQ-----ERQYEEELQOEEBQLLEER--EKR 968  
 QY 1194 RWRATIQMRERKNYLHQTTKRI-----QIKFRAKREMKQARFLOLKVTLVVOX 1248  
 DB 969 RQREERQYKDKK----LQKEQLLGEPEKRRQERKRYKEELQOEEBQLLEER 1024  
 QY 1249 RRALLQMRKERQBYLHLREVTILQRRFHAQKSMRPMRAKYRGTOAVSLQWHPNHL 1308  
 DB 1025 RE-----KRRQBW--ERQYKDKDELQOEEBQLLEERKRR-----LQRRERQY- 1067  
 QY 1309 LRKRERNFLQRAAITLQRRYARLNMILKLSYALQKQAITI-----QTRYAKKA 1363  
 DB 1068 ---REEELQOEEBQLLEERETRRQERLERQYRKKEELQOEEBQLLEERPEKRRQERE 1124  
 QY 1364 MQ-KQVLYQKORAITKQRYVGNLEMRKQI---EVYQKQOAVIRLQ-----KW- 1411  
 DB 1125 RQREELQOEEBQLLEERKRRQERLERQYRKKEELQOEEBQLLEERQYRDEQSDLKWQ 1184  
 QY 1412 W-----RSIRDMRL-CKA-----GYRRI-----RLSSLSTIQ-----RKW 1439  
 DB 1185 WEPEKENAVRDNKYCKRGNEQFQLEDQSVDRQSQDQLHLLGQEQERDRBQRRRW 1244  
 QY 1440 RATVQARR-----QREIFLSTIRKVLQMOPFIRATILMQORRFEMKRRAAVQIR 1491  
 DB 1245 Q---QANRHPFEEQLEERQEKAKRDRKSQ---EBQQLLEEREE---KRRRQET-DR 1294  
 QY 1492 RFRARCAMLKARQDYQLIQSVILVQKFRANKSM--KQARQEVQVLRITAVHLQOKF-- 1547  
 DB 1295 KFRREEQLQEREEQPLLRQE---RDKKFRFEEELHQQGKFLFEEQRLEERERFLK 1351  
 QY 1548 RGKRLMTEQRNCFQLRCSMPGFGQARAGFMARKRFOALMTPMMDLIRQKRAKVIQRY 1607  
 DB 1352 EQQLRLER---EQLR-----QDRDKF-----REEEQQLSRQERDKFRREE 1392  
 QY 1608 WRGVLTERRQKHQGLDIRKRIALQROE 1635  
 DB 1393 QQ---VARQRRERKFLLEEQ---QLRQE 1414

RESULT 15  
 ADD48869  
 ID ADD48869 standard; protein; 1898 AA.

XX AC ADD48869;  
 XX DT 29-JAN-2004 (first entry)  
 XX DE Human Protein AAA65582, SEQ ID NO 14580.  
 XX KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX OS Homo sapiens.  
 XX PN W02003016475-A2.  
 XX PD 27-FEB-2003.  
 XX PF 14-AUG-2002; 2002WO-US025765.  
 XX PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR GENBANK; AAA65582.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 PS Claim 1; Page; 1017pp; English.  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1898 AA;  
 Query Match 2.9%; Score 277; DB 7; Length 1898;  
 Best Local Similarity 21.7%; Pred. No. 3.2e-11;  
 Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

QY 891 RRHLVHVVIQRIIPKELMRHRAATVIOAFVGHOMKVKVLFKTFRTQAAILQKFTF 950  
 DB 648 RRHLLKSEQERHRLERHRAATVIOAFVGHOMKVKVLFKTFRTQAAILQKFTF 950

```

QY 951 RYLAQKQLYQSYHSIIT-IQW-WR-----AQOLG----- 978
Db :||: :|: :|: :|:
QY 702 QELAEERQEQARERIKSIPQWQWQSEADARQSKVILEAPQAGRAPOEQEKKRRRE 761
Db :||: :|: :|: :|:
QY 979 -----RQHRQFVLEAAIFLQIW-----RRRLFAKLJAAAEARLQRS 1020
Db :||: :|: :|: :|:
QY 762 SELQWQEEERARQOQEEQERDFTWQ--WQAEKSEGRQOELSARPPFLRQERQOLRAE 819
Db :||: :|: :|: :|:
QY 1021 KQQAASYYIQMWRTYOLGRIORH-----BFLQORDLIMFVQRRMRKSNMLEQ--RKE 1073
Db :||: :|: :|: :|:
QY 820 ERQOREQFLPEEBEKEQGRQRREREXELQFLEEEQOL--QRRERAQOLQEEEDGLQE 876
Db :||: :|: :|: :|:
QY 1074 FQOLKRAINITQORWAKLSMRKCNADYALRSSLVLYQAYRKATIQMRIDENHYISLRK 1133
Db :||: :|: :|: :|:
QY 877 DOERRQORQDQKRWQLEBERKRRH-----TLYAKPALQEQ-----LRK 918
Db :||: :|: :|: :|:
QY 1134 NVICIQORLRAIMKMRQENYRLRNASIIIVQKRYMRQMIQDRNAVYLRTRKCIINVO 1193
Db :||: :|: :|: :|:
QY 919 EQOLQEEEBEL--QREERKRRQEQ-----ERQYREEQOLQEEEQOLLEER---EKR 968
Db :||: :|: :|: :|:
QY 1194 RRWRATLOMRERKNYLHLOTTTKRI-----QIKPRAKREMKQRAEFLOLKKVILVVOK 1248
Db :||: :|: :|: :|:
QY 969 RRQERQYRKDKK---LQKEEQOLGEEPEKRRQERKRYREEBELQOEEEQOLLEE 1024
Db :||: :|: :|: :|:
QY 1249 RRRALLQMRKREQVYLHLREVTIKLORRFHQAQSMRFMRKRYRGTAQAAVSCLOMHRNHL 1308
Db :||: :|: :|: :|:
QY 1025 RE-----KRRQEW--BROVKKDELQOEBEQOLLEERKRR-----LQERERQY- 1067
Db :||: :|: :|: :|:
QY 1309 LRKRENSFLQROAAITLQRRYARLARLNMKOLKSYAOLKQAAITI-----QTRYRAKKA 1363
Db :||: :|: :|: :|:
QY 1068 ---REEELQOEEEQOLGEEERETRERQELERQYRKEBELQOEEEQOLLEERPEKRRQERE 1124
Db :||: :|: :|: :|:
QY 1364 MQ-KQVVIYQKOREAIKVQRRYRGNLENMKOI-----EYVQKROAVIRLQ-----KW- 1411
Db :||: :|: :|: :|:
QY 1125 RQCREEBELQOEEEQOLLEERKRRQELERQYREEBELQOKRQRYRDEQDSDLKWQ 1184
Db :||: :|: :|: :|:
QY 1412 W-----RSIRDMRL-CKA---GYERI-----PLSSLSIQ-----EKW 1439
Db :||: :|: :|: :|:
QY 1185 WEPEKENAVRDNKVYCKGRENQFQHLEDSQVRDQSQODLQHLIGEQQOERDQERRRW 1244
Db :||: :|: :|: :|:
QY 1440 RATVQARR-----QREIFLSTIRKVLMOAFTRATILMRQORREFEMKRRAAVVIQR 1491
Db :||: :|: :|: :|:
QY 1245 Q---QANRHFPPEEQOLEREEQEKAKRRDKSQ---EEKQLLEEREE---KERRQET-DR 1294
Db :||: :|: :|: :|:
QY 1492 RFRARCAMLKARQDYQLIQSSVILVQKFRANRSM--KQARQEFVOLRTIAVHLOQKF-- 1547
Db :||: :|: :|: :|:
QY 1295 KPREEBEQLEEREQOLLRQE---RDRKPREBELLHQGRKFLBEEQOLREERERKFLK 1351
Db :||: :|: :|: :|:
QY 1548 RKGRLMIQRCNCFQLLRCMPGQFARARCFMARKRFQALMTPEMMDLIROKAAKVIQRY 1607
Db :||: :|: :|: :|:
QY 1352 BEQQLRLEBER---EQIR-----QDRDKF-----REEEQOLSRQERDRKFRREE 1392
Db :||: :|: :|: :|:
QY 1608 WRGYLIRRRQKHQGLLDIRKRIAQLRQE 1635
Db :||: :|: :|: :|:
QY 1393 QQ---VRQERERKFLBEEQ---QLRQE 1414
Db :||: :|: :|: :|:

```

Search completed: September 13, 2004, 17:40:45  
Job time : 117 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:37:54 ; Search time 36 Seconds  
(without alignments)  
4972.567 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: Piri:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9514	100.0	1861	T13845	microtubule-associ
2	361.5	3.8	1186	T19957	hypothetical prote
3	324.5	3.4	1088	T05464	hypothetical prote
4	324	3.4	1407	S28589	trichohyalin - rab
5	323.5	3.4	1549	A40691	trichohyalin - she
6	320	3.4	1830	S19188	myosin-V - chicken
7	314	3.3	1909	A45592	liver stage antigen
8	304	3.2	1828	B59254	myosin heavy chain
9	301	3.2	4574	G02520	myosin heavy chain
10	301	3.2	4687	A38638	plectin - human
11	300.5	3.2	1375	T05200	plectin - rat
12	300.5	3.2	1855	A59254	myosin heavy chain
13	300	3.2	1374	D85390	myosin heavy chain
14	294	3.1	4684	A59404	myosin-like protei
15	284.5	3.0	1805	A64224	plectin [imported]
16	277	2.9	1818	S73852	hypothetical prote
17	277	2.9	1898	A45973	hypothetical prote
18	271.5	2.9	1853	A46761	trichohyalin - hum
19	267.5	2.8	2663	S28261	myosin heavy chain
20	264	2.8	2442	T08621	centromere protein
21	260	2.7	2245	T18278	centrosome associ
22	254.5	2.7	1837	T41023	myosin heavy chain
23	252	2.6	1313	A48467	probable nuclear p
24	251.5	2.6	1792	T13939	myosin heavy chain
25	250	2.6	2101	A42184	myosin V - fruit f
26	246.5	2.6	2022	T43214	nuclear mitotic ap
27	244	2.6	1846	A59289	ovtl protein - nem
28	243.5	2.6	812	A53016	myr 6, unconvectio
29	243.5	2.6	1410	A57013	myosin heavy chain
					early endosome ant

## ALIGNMENTS

### RESULT 1

T13845

microtubule-associated protein - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000

C;Accession: T13845

R;Saunders, R.D.; Avides, M.C.; Howard, T.; Gonzalez, C.; Glover, D.M.

J. Cell Biol. 137, 881-890, 1997

A;Title: The Drosophila gene abnormal spindle encodes a microtubule-associated protein t

A;Reference number: Z17792; MUID:97296495; PMID:9151690

A;Accession: T13845

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1861 <SAU>

A;Cross-references: EMBL:U95171; NID:g1930121; PID:g1930122; PIDN:AAB51540.1

C;Genetics:

A;Gene: asp

A;Cross-references: FlyBase:FBgn0000140

C;Function:

A;Description: is required for the normal function of the mitotic spindle

Query Match	100.0%;	Score 9514;	DB 2;	Length 1861;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1861;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELVWSPVLEVACKETLQLIDNFRKEVWILKSKNOPVKNPKFPVTGKTLQKSP	60	
Db	1	MELVWSPVLEVACKETLQLIDNFRKEVWILKSKNOPVKNPKFPVTGKTLQKSP	60	
QY	61	GAGTKMKSIVSAAVQKKRMSAAAAAPPQKQTVRTAPSRPAWAHPPQAPLVEKNVKT	120	
Db	61	GAGTKMKSIVSAAVQKKRMSAAAAAPPQKQTVRTAPSRPAWAHPPQAPLVEKNVKT	120	
QY	121	POEEPVYLSPOPSLKENLSPMTCNLIDVINDLFTPLTETRGKQATIPDNLAAPT	180	
Db	121	POEEPVYLSPOPSLKENLSPMTCNLIDVINDLFTPLTETRGKQATIPDNLAAPT	180	
QY	181	PTLKNVKSANDMEPRITPDDLEDQATNKTFDVHSETINISLDTDCSRIDGQPH	240	
Db	181	PTLKNVKSANDMEPRITPDDLEDQATNKTFDVHSETINISLDTDCSRIDGQPH	240	
QY	241	PLNKTITTVHATHRALACIHEERGSPRPTPTKSAIHDLRDKIKVGLSPRKYS	300	
Db	241	PLNKTITTVHATHRALACIHEERGSPRPTPTKSAIHDLRDKIKVGLSPRKYS	300	
QY	301	LSLLSPQTKYAIQSGMPNLNEMKIRSIQNYQEQQIQIKAKDLNLSSSSSSEAS	360	
Db	301	LSLLSPQTKYAIQSGMPNLNEMKIRSIQNYQEQQIQIKAKDLNLSSSSSSEAS	360	
QY	361	FLFNHSEILAQSSRFNLHEVGKSVKGVKPNPKRRSHLSFSDAPSNESIYRN	420	
Db	361	FLFNHSEILAQSSRFNLHEVGKSVKGVKPNPKRRSHLSFSDAPSNESIYRN	420	

QY 421 SPKQKQVEDTTLPRSAAPANASARSSSAHAWPHAQSKFKPLAQTWMSLMKKPATPRKVRD 480  
 DB 421 SPKQKQVEDTTLPRSAAPANASARSSSAHAWPHAQSKFKPLAQTWMSLMKKPATPRKVRD 480  
 QY 481 TSIPQSVKLYDSELYMOTCINPDPAATTTIDPELASTMYLDEQAVDRHQADPKWLNAL 540  
 DB 481 TSIPQSVKLYDSELYMOTCINPDPAATTTIDPELASTMYLDEQAVDRHQADPKWLNAL 540  
 QY 541 VSPADLADLNNKIDVGLFNEVRNKLIVAPTKEEQSMNYLTIVRLETTLKAQAVELFF 600  
 DB 541 VSPADLADLNNKIDVGLFNEVRNKLIVAPTKEEQSMNYLTIVRLETTLKAQAVELFF 600  
 QY 601 SQOMRLPCSKVAVYNNKQALRIRSDENHLVDVMORTILELLCNPLWRLGLGVVGE 660  
 DB 601 SQOMRLPCSKVAVYNNKQALRIRSDENHLVDVMORTILELLCNPLWRLGLGVVGE 660  
 QY 661 KIOMQSNRDI VGLSTFILNRLFNKCEBQYSKAVTLTEEYAEITIKHSLQKILFLLPFL 720  
 DB 661 KIOMQSNRDI VGLSTFILNRLFNKCEBQYSKAVTLTEEYAEITIKHSLQKILFLLPFL 720  
 QY 721 DOAKQKRIYKHNPCLVKSPKPKETKDILLRSESSLLANIGDITRELRLGHVLOHROT 780  
 DB 721 DOAKQKRIYKHNPCLVKSPKPKETKDILLRSESSLLANIGDITRELRLGHVLOHROT 780  
 QY 781 LDEFYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPAISRLQRIFNVKLALGALG 840  
 DB 781 LDEFYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPAISRLQRIFNVKLALGALG 840  
 QY 841 EANFOLGGDIAAQDIDVGHREKTLISLMOLIYKFRSPKPHAAATVLOKWRHMLHVVIQ 900  
 DB 841 EANFOLGGDIAAQDIDVGHREKTLISLMOLIYKFRSPKPHAAATVLOKWRHMLHVVIQ 900  
 QY 901 RRIKHELMRRRAATVIOVFRGHOMRKVYKLFKTERTOAAIILOKETRYLAQKOLYO 960  
 DB 901 RRIKHELMRRRAATVIOVFRGHOMRKVYKLFKTERTOAAIILOKETRYLAQKOLYO 960  
 QY 961 SYHSIITIQWRRAAQOLGROHQRVVELREAAIFLQRIWRRRLFAKLLAAETARLORS 1020  
 DB 961 SYHSIITIQWRRAAQOLGROHQRVVELREAAIFLQRIWRRRLFAKLLAAETARLORS 1020  
 QY 1021 OKQOAAAASYIQOWRTYQIGRTORHEFFLORDLIMFVORRMSKWSMLEQKEFQOLKRA 1080  
 DB 1021 OKQOAAAASYIQOWRTYQIGRTORHEFFLORDLIMFVORRMSKWSMLEQKEFQOLKRA 1080  
 QY 1081 AINIQQWRRAKLSMRKCNADYLALSSVLKVOYKATIQMRIDRNHYISLRKNVICLOQ 1140  
 DB 1081 AINIQQWRRAKLSMRKCNADYLALSSVLKVOYKATIQMRIDRNHYISLRKNVICLOQ 1140  
 QY 1141 RIRATMKQREORENLYRLNNAISILVQRYMRQMIQDRNAYLRTKCIINVORRWATL 1200  
 DB 1141 RIRATMKQREORENLYRLNNAISILVQRYMRQMIQDRNAYLRTKCIINVORRWATL 1200  
 QY 1201 QMRERKNYLHLOTTTKRIQIFRAKREMKQBAEFLOLKKVTLVQKRRALLQMRKER 1260  
 DB 1201 QMRERKNYLHLOTTTKRIQIFRAKREMKQBAEFLOLKKVTLVQKRRALLQMRKER 1260  
 QY 1261 QBYLHLREVTIKLORRFAHQKSMRPMKAKYRGTOAAVSCLOQWNRHLLRKEERNSFLOL 1320  
 DB 1261 QBYLHLREVTIKLORRFAHQKSMRPMKAKYRGTOAAVSCLOQWNRHLLRKEERNSFLOL 1320  
 QY 1321 RQAATLQRYRVARLNNIKOLKSYAQLOKQAAITITOTRYRAKMKOKOVLYOKOREALIK 1380  
 DB 1321 RQAATLQRYRVARLNNIKOLKSYAQLOKQAAITITOTRYRAKMKOKOVLYOKOREALIK 1380  
 QY 1381 VQRRYRGNLEMRKQIEVYQOKQAVIRLQKWRWSIRDMLKAGYRIRLSSLSIORKWR 1440  
 DB 1381 VQRRYRGNLEMRKQIEVYQOKQAVIRLQKWRWSIRDMLKAGYRIRLSSLSIORKWR 1440  
 QY 1441 ATVQARROREIFLSTIRKVRMLMOATIRATILMROORREFEMKRAAAVIOQRFFARCAML 1500  
 DB 1441 ATVQARROREIFLSTIRKVRMLMOATIRATILMROORREFEMKRAAAVIOQRFFARCAML 1500

QY 1501 KARQDYQLIOSSVILVORKFRANRSMKQARQEFVQLRTIAVHLQCKFRGKRLMIEQRNCF 1560  
 DB 1501 KARQDYQLIOSSVILVORKFRANRSMKQARQEFVQLRTIAVHLQCKFRGKRLMIEQRNCF 1560  
 QY 1561 QLLRCSMPGFOARARGFMARKRFOALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRKQKHQ 1620  
 DB 1561 QLLRCSMPGFOARARGFMARKRFOALMTPEMMDLIRQKRAAKVIQRYWRGYLIRRKQKHQ 1620  
 QY 1621 GLLDIRKRIAQLRQEAQKAVNSVRCKQEAQVRFRLGRFIAASDALAVLSQDLRLSRTVPHLL 1680  
 DB 1621 GLLDIRKRIAQLRQEAQKAVNSVRCKQEAQVRFRLGRFIAASDALAVLSQDLRLSRTVPHLL 1680  
 QY 1681 MWSEFMSTCYGIMQAQIRSEVDKQIJCPCRSIILNLARYNSTVNTFOEGGLVNTIAQM 1740  
 DB 1681 MWSEFMSTCYGIMQAQIRSEVDKQIJCPCRSIILNLARYNSTVNTFOEGGLVNTIAQM 1740  
 QY 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKRRKIIHDYMTNPEAIYVRETKKLVARKEKMQ 1800  
 DB 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKRRKIIHDYMTNPEAIYVRETKKLVARKEKMQ 1800  
 QY 1801 NARKPPMTSGRYKSKINFTPCSLPSLEPFGIIRYSPYTFISSVYAFDTILCKLQIDM 1860  
 DB 1801 NARKPPMTSGRYKSKINFTPCSLPSLEPFGIIRYSPYTFISSVYAFDTILCKLQIDM 1860  
 QY 1861 F 1861  
 DB 1861 F 1861

RESULT 2  
 T19957  
 hypothetical protein C45G3.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T19957  
 R:Barlow, K.  
 submitted to the EMBL Data Library, March 1997  
 A:Reference number: Z19203  
 A:Accession: T19957  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1186 <N1>  
 A:Cross-references: EMBL:Z92780; PIDN:CAB07174.1; GSPDB:GN00019; CESP:C45G3.1  
 A:Experimental source: clone C45G3  
 C:Genetics:  
 A:Gene: CESP:C45G3.1  
 A:Map position: 1  
 A:Introns: 21/1; 255/2; 363/2; 575/3; 893/3; 1017/2; 1042/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C45G3.1

Query Match 3.8%; Score 361.5; DB 2; Length 1186;  
 Best Local Similarity 18.6%; Pred. No. 4.1e-12;  
 Matches 244; Conservative 216; Mismatches 480; Indels 375; Gaps 47;

QY 458 KKPFLAQWLSMKK---PATPKVRVTSIQPSVKLYDSLYMOTCINPDPAATTTIDPF 514  
 DB 21 EKLLDQVKSNTKKIDLRATERAFLESS-----PTSMNMTPLNPS-ISSVSQDSFI 71  
 QY 515 LASTWYLDQAVDRHQADPKKMLNALVSI PADLADLNNKIDVGLFNEV----- 564  
 DB 72 LS-----YDEKA-NKQIIALATWCNTM-----NELDVSSEMDLGSKACRNIQWMLKK 120  
 QY 565 RNKELVVAPTKEQSNY---LTKYRLETLIRKAAVELFFSEQMRILPCKSVAVYNNKQALR 621  
 DB 121 RSTSEVENTOENARRRYQRIFEKNDPEVVKCKCKQLDDSGMD---ASIKDLSKNVYA 177  
 QY 622 IRSDRNHLVDVMORTILELLCNPLWRLGLGVVGEKIQWQSNRDI VGLSTFILNEL 681  
 DB 178 IRKEHAVYNDIGLQTLILHTFFSFHPANLKTALAEALFNTRIDAPKHLMKLSQFFLDLV 237  
 QY 682 FRN--KCEBQYRSKAY---TLTEEYAEITIKHSLQKILFLLPDLQAKQKRIYKHNPCLF 736  
 DB 238 FSNPTLKNKKAQAGSKPDIITTEAGKALHGHFLSVSMKLMFLIETIHTHVRPILNLTIF 297





R.Fietz, M.J.; McLaughlan, C.J.; Campbell, M.T.; Rogers, G.E.  
 J. Cell Biol. 121, 855-865, 1993  
 A>Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding  
 A:Reference number: A40691; MUID:93260018; PMID:7684041  
 A:Accession: A40691  
 A:Molecule type: DNA  
 A:Residues: 1-1549 <FT2>  
 A:Cross-references: EMBL:218361; NID:g295940; PIDN:CAA79165.1; PID:g295941  
 A>Note: sequence extracted from NCBI backbone (NCBI:132511)  
 R.Fietz, M.J.; Presland, R.B.; Rogers, G.E.  
 J. Cell Biol. 110, 427-436, 1990  
 A>Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker  
 A:Reference number: A34209; MUID:90130632; PMID:2298812  
 A:Accession: A34209  
 A:Molecule type: mRNA  
 A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <FT2>  
 A:Cross-references: GB:X51695; NID:g1827; PIDN:CA35992.1; PID:g1828  
 C:Comment: trichohyalin is a protein of the medulla of the hair and of the inner root sheath  
 Covalent modifications to this protein include conversion of arginine to citrulline and  
 C:Genetics:  
 A:Introns: 46/3  
 A>Note: single copy gene  
 C:Superfamily: trichohyalin; calmodulin repeat homology  
 C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat  
 F:49-81/Domain: calmodulin repeat homology <EF2>  
 F:387-851/Region: 28-residue repeats  
 F:886-1519/Region: 23-residue repeats

Query Match 3.4%; Score 323.5; DB 1; Length 1549;  
 Best Local Similarity 19.3%; Pred. No. 7.6e-10;  
 Matches 280; Conservative 249; Mismatches 504; Indels 421; Gaps 60;

QY 282 RDIKLVGSLPKYSBMSK-DLSLLSPQTKYATOGSMENLNE-----M 322  
 DB 308 RQKYSPRHQEQSRQEQELLERQEQQLSEVQSLEQDQGRQLKQEQRYDQNRW 367  
 QY 323 KIRSTEQNRY-----QEQTIQKADLNSSS-----SEASLAGQEQEF 361  
 DB 368 QLEESQRRYTYAKPAQEQVREEQQLRLKEKLQKREKQERQYREVELQBER 427  
 QY 362 LFNHSEIIAQSRFNLHVEGRKSVGSPVKNPKRSRH-----LSFSDAPSNESLYRNE 416  
 DB 428 LQREBEQQLQREER-----EKRRQREKQYLEKVELMEEQQLQREE 468  
 QY 417 TVAISPPK-KQKVEDTTLPSAAPANASARSSAHAWHAQSKPKLAQTWSLMKKPATP 475  
 DB 469 REKQREKQYLEKVEL-REBEQQLQREK-----RQREKQYLEKVELQEQEQ 521  
 QY 476 RKVDRTSIQPSVKLYDSLYMOTCINPDPAATTTIDPFLASTMYLDQAVDRHQADPKK 535  
 DB 522 REERKRRQREKQYLEKVELQ-----BEEQLQREKREK 557  
 QY 536 WINALVSTPADLDLNNKIDV-----GKLFNEVNTVELVAPTKEQSNMYLTKYRL---E 589  
 DB 558 -----QEREKQYLEKVELQEQEQQLQREKQK-----RQREKQYLEKVELQEQ 602  
 QY 590 TLRKAVALFFSEQMLPCSKVAVVNVKQALIRSDR-----NLHLDVVMQRTILEL-- 641  
 DB 603 QIQREKREKQREKQYLEKVELQEQEQVQREKQREKQREKQYLEKQLQBERLOE 662  
 QY 642 ---LLCFNPLMLRLGLEVVFGEKIQMQQNRDIVGLSTFILNLFNFKCQEQRYKAYTLT 698  
 DB 663 BEQLLREERKRRQREKQYLEKVELQEQEQ-----LQREERKREKQER-BEQYLEK 713  
 QY 699 EBYAETIKKSLKILFLPDLQAKQKRIYKHNPNCLFVKSPKHEKDI-----LLRFS 753  
 DB 714 BELQK-QEELQREKQEQLEQREKQREKQVREK---YLEEELQEQEDRLQREKQLRED 768  
 QY 754 S-----ELLANIGDITRELRGLVYLOHQRTFLDFEYAFNNLAVDLRDGVLTFRV 804  
 DB 769 REKQYLEKVELQEQEQQLQREKQREKQYREEL-----LREERLHK 815  
 QY 805 VEVI-----LLRDLRLQLRVFAISRQLRIFNVKLAGALGANF-----QLGQDIAAQ 853

DB 816 EQQLQREBEKRRQRELEQLQEEELQRLDR-----KROFRDDQHQNEVRNS 863  
 QY 854 DIVDGHRE-----KTLSSLWLIYKPRSPKFFHAAATVLQ-----KWMR-----R 892  
 DB 864 RVYSKHRENKESRLQDDSW-----VRESQFQDLRPLQDEQEKREKREKQERKQKDS 918  
 QY 893 HMLVVIQRIHKEMLMRHRAATVIQAVFRGHWRK--YV-----KLFKTEET 939  
 DB 919 QPFAQLLEREQOKETERDRKREBEQQLKQREKIRYLEEDRKFRFEQQLRELE 978  
 QY 940 Q-----AAITLQKFTRRYLAQKLYQSYHSIIITQWRWRAQQLGRHQKQ--FVE 987  
 DB 979 QQLQREDRKFRFEELSQRQERDRKFRFEQQLQEQ-----REQRURQERDRKFRFE 1027  
 QY 988 -----LREAAIFLQRIWRRLP-----AKKLLAAAFARLQSQKQAAASYIQMOW 1034  
 DB 1028 ERQLQREBEQRLRQERDRKFRFEQQLQEQREBEQRLRQERDRKFRFEQQLLREEQQL 1087  
 QY 1035 RYVQLGRIGRHEFLQRDLIMFVQ-----RRMSKWSMLEQRKEFOQLKRAAIN 1083  
 DB 1088 ROERNRKRFEELRERBEQQLQEQGEQQLQEQGEQQLQEQGEQQLQEQGEQQLQEQGE 1145  
 QY 1084 IQORWRKLI-----SMRKCND-----YLALRSSVLKQVAYRKATIQMRIDRNHY 1129  
 DB 1146 RKFREBAQILKEBEQRLRQERDRKFRFEQQLQEQEELR-----QEREPQLQERDRKF 1201  
 QY 1130 SURKNVICLOQLRAIMKQREORENYLRNNSIIVQKRYMRQOMIQDNAYLTRKCI 1189  
 DB 1202 REEQQLQREKRL-----RQREPOLQER-----DRKFHEEQQLQREBEQQLRQEQ-- 1249  
 QY 1190 INVQERWATLQMRERKNYLHLQTTXRIQIKFPAKEMKQBAEFLOKVVILVQKR 1249  
 DB 1250 ---RDKFREBAQILQREBEQRLRQERDR-----KFRFEQQLQEQEELR-----QER 1296  
 QY 1250 PRALL-----QMRKEROYLHLREVITIKLQRRHAQ-----KSMRFWRKRYGTGQAASVCLQ 1301  
 DB 1297 DKFREBEQQLQREBEQQLRQEQ-----RDKFREBEQQLKEBEQQLRQERDR-----K 1345  
 QY 1302 MHWNRHLKREKERNFLOLRQAATFLORRYRARLNMKQLKSYAQQLKAAITIQTRYAK 1361  
 DB 1346 PHEKELLRLREEQ--QLRQELE-----GVFSQEQELRAEQEBEQERQQRDR 1393  
 QY 1362 KAMQKQVLYQKREAIIVQRRYRGVLEMKQTEVYQKQAVIRLQKWRWSIRDMRLC 1421  
 DB 1394 KPLEEQQLQREBEKREKRVQEQDKKFLQEQEQ-----HREBEQELRRQO----- 1440  
 QY 1422 KAGYRRIRLSSISIQKWRATVQARRQREIFLSTIRKVLQMAFTRATLLMRQORREFM 1481  
 DB 1441 -----LDQYRAEQFAE-----EKRRQEQELQEQEQRQRE----- 1476  
 QY 1482 KRAAVIQRRFRACAMKAKQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAV 1541  
 DB 1477 -----RKFRBEQQLRQEQEQ-----EQKRRQERDVQOSRRQ----- 1507  
 QY 1542 HLQKFKRGKRLMIE 1555  
 DB 1508 -VWEEDKGRQVLE 1520

## RESULT 6

S19188

myosin-V - chicken

NAlternate names: calmodulin-binding protein; myosin I heavy chain, 190K

C:Species: Gallus gallus (chicken)

C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text\_change 19-Jan-2001

C:Accession: S19188; A44359; B44359; S29249

R:Espresso, E.M.; Cheney, R.E.; Matteoli, M.; Nascimben, A.A.; De-Camilli, P.V.; Larsc submitted to the EMBL Data Library, February 1992

A:Description: Complete cDNA coding sequence of chicken brain p190, a calmodulin binding

A:Reference number: S19188

A:Accession: S19188

A:Molecule type: mRNA









1178 DRNAVLRTRKCIINVQREWRATLQWREERKKNVHLQTTTKRIQIKFR----- 1224  
1048 QAKEMTETMEKKL-VEETKQLELDLNDRLRYQNLNFSLEERYDYLKBEWTLWHPV 1106  
1225 -----AKRMKKQRAEFLQLKKVT-----LVQKERRALL 1254  
1107 KPGHKRTDTHSSNESEYIFSSIEAEMEDIPRTEEPSEKKVPDLSLFLKQKRVTELE 1166  
1255 QMRKERQEYHLHREVITKLRFFHQAQKMRFWAKYRGTQAQVSCLOMHWNRHL--LRKR 1312  
1167 QEKQVMDLDRKBEQVL-----RSKAKEBERPOIRGALEYESLK--RQLESENKK 1217  
1313 ERNSFLQRLQRAAITLQRR-----YRRLNMWKQLKSYAQLKQAAILTQTRVA 1360  
1218 LKNEINELRKA---LSEKSAPEVTAPGAPAYRLMEQLTSVSEELDVRKEEVL1---LRS 1271  
1361 KQAMQKQVLYQKOREAI1-----KVQRRYRGNLEMKKQLEVTYQKQRA 1404  
1272 QLVSQEAKIAPKDDKNTMTDSTILLEDVQKMDKGEINAQYIQLKETRWLE----- 1323  
1405 VIRLOKWRSRD-----MRLCKAGYRIRILSSLSIQKWRATVQARRQREIFLST 1455  
1324 -SLOQSKRSHENAEALRGEIQSLKENNRQQOILLAQNQLPPEARTEASLQHEITRLT 1382  
1456 IRKVRLMQAFTRATLLMRQORREPEMKRAAVIQRFRARCAMLKARQDYOLIQSSV-- 1513  
1383 NENLDLMEQ-----LEKQDKTVRKLKQLKFAKKI GELEVGOMENISPGOIIDEP 1436  
1514 ILVORKFRANRSMKQARQEFVQLRTIAVHLQKQFGRKELMIEQRNCFOLLRCSPGFOAR 1573  
1437 VNI PRKEKDFQGMLEYKKEDEQKLVKNLILLEKPRGAVNL-----LPGLPAY 1484  
1574 ARGFMAR-----KRFQALMTFEMMDLIRQKRAAKVIQR-----YWRGYLIR-- 1614  
1485 ILFCVVRHADYLNDDQKVRSLTSTINS1-----KKVLKKRGDDFETVSFWLSNTCRFL 1538  
1615 -----RROKHQGLLDIRKRIALROEAKAVNSVRCKV-QEAVRFLRG- 1655  
1539 HCLKQYSGEGBGFMKHTNSRQNEHCLTNF--DLAEYRQ---VLSDLAIOIYQQLVRVLENI 1593  
1656 --RFIASDAL-----AVLSQLDRLSRTPVPHLLMCMSEBFMSTFCYGIM 1695  
1594 LQPMIVSGMLEHETIQGVGKPTGLKRTSSIADEGTYTLDSILRLQNSFHSVMCOHGM 1653  
1696 AQAIRSEVDKOLTERCSRILN-----LARYNSTTV-----NTFOEGG 1733  
1654 DPBLIKQVVKOMFYIIGAITLNNLLLFKDCMSWCKGMQIRYNVSQLBEWLDRKXLMNSGA 1713  
1734 -----LVTTIAQML--LRWCDKDSIEINTLCTLI 1759  
1714 KETLEPLIQRAQLLVQVKKTDDEAELCSMCNAL 1747

RESULT 9  
G02520  
plectin - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
C:Accession: G02520  
R:McLean, W.H.I.; Smith, F.J.D.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: H01385  
A:Accession: G02520  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4574 <MCL>  
A:Cross-references: EMBL:U53204; NID:g1477645; PIDN:AAB05427.1; PID:g1477646  
C:Genetics:  
A:Gene: PLECL1  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S  
F;68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

Query Match	3.2%;	Score 301;	DB 2;	Length 4574;
Best Local Similarity	19.6%;	Pred. No. 6.1e-08;		
Matches 334;	Conservative 276;	Mismatches 652;	Indels 446;	Gaps 66;
QY	161	ETRGKGQATIF--PDNLAAMP-----TPTLKGNVKS	CANDMRPRITPD	DLEDQDPATNKT 213
DB	1061	EGLGKGVARLSAAEKVIALPEPSPAATTURSELE-----	LTGLKLEQVRSLSAI 1110	
QY	214	FDVKHSETINISLDTDCSRIDGPHPTPLNKTITTVIHATHTR--	ALACIHEEBCPPPR 270	
DB	1111	Y-LEKLKTIISIVIRGTQGAEBVLRAHEEQLEKAQAVPATLPELEATV	KASLKLRQAQAEQ 1169	
QY	271	TPTKSAIHDLKRDIKLVGSPL-RKYSSEMKDLSLLSPQTKVIAIQGSM	PNINEMKIRSI-E 328	
DB	1170	QPTFDALRELRGAQGEVGERLQORGHGRDLEVERWRERVAQLMER	QVLAQTDLRQREL 1229	
QY	329	-----QNRYYOE-----QOIQIKADLNSSSSSEASLAGOO--	EFLFNHS 366	
DB	1230	EOLGRQLRYRESADPLGAWIQDARRQEQIQAMPLADSOAVREQLRQ	EOQALLIEERHG 1289	
QY	367	FILAQSGRF-----NLHVEGRKSVKSGPKVKNPHKR-----	RS 398	
DB	1290	EKVEECQRFQAKYINAIKDVELQLVYTKA-QLPEVPASPAKPKVQS	GSSESVIOEYVDLRI 1348	
QY	399	HELFSOAPS-----NESLYRNETVAISPKKQORVEDTTLPRSAAP	ANASARSSSAHAW 452	
DB	1349	HYSLTITLTISQYIKFISETLRMBEERLAEQORAERERLAEV	EALAEKQRLAEHAQ 1408	
QY	453	PHAQSKFKLAQWLSMKKATPRKVDRTSIOPSVKLYDSEL-YMQTC	INPDPFAATTYI 511	
DB	1409	AKAQER-EAKELQORMOEYVVRREEAADAQOQKKSIOEELQQLRQ	SSAEITQAKARQA 1467	
QY	512	DPFLASTWYLDE-----QAVDRHQADFKKWLNALVSIIPADLAD	LNN-KIDVGKLF 561	
DB	1468	EAASRLRIIEEIRVYRLQLEATERQCGAKSELQALRAEAEAEAKRQ	AGEAERLR 1527	
QY	562	NEV-----RNKELVVAPTKEBOSMYLTKYRLLET-----	LKKAAYELF 599	
DB	1528	QVQDESQRKQAEVELASRVKAETEAREKQALQALBELRLQAEAE	ERRLRQAEVERA 1587	
QY	600	FSEQWRLPC-----SKVAVYVYKQALRTRSDRNHLDVDVMORT	ILELLLCFNPLW 649	
DB	1588	RQVQVALETQRSABAELOSRAFSFAKTAQLERSLQEEHVAVQAUR	EEAE-----1638	
QY	650	LRLGLEVVVFGEKIQMOSNRDIDVGLSTFIL--NRLPRNKCEORYS	KAYTLTBEYAEITKK 707	
DB	1639	-RRAQQQAEAREAEAREE--LERWQLKANEARLRLQAEVVAQKSLA	QABAE-----1690	
QY	708	HSLOKILFLPFDIQAKQKRIKVRNPCLFVKYSPHKETKDILLRFS	SELLANIGDITREL 767	
DB	1691	-----KQKE-----	BAEAA 1700	
QY	768	RELGYVLQH--RQTFELDEFYAFNNLAVDLRQVRLTRVVVEVILLR	-----DDLTRQLRV 820	
DB	1701	RRKGABEQAVRQELABAELEKQORQALAEQTAQORLAAEQELIRL	RAETEQEGEQOQLLE 1760	
QY	821	PAISRQLQRIFNVKALCALGEANFOLGGDIAAQDIDVGHREKTL	SILLWLIYKFSPPKH 880	
DB	1761	EELARLQ-----EAAAATQKQELAEALA-----KVRAE--	1790	
QY	881	AAATVILQKWRHHLHVVIQRRIRHKELMRHRAATVIOAVFRGHQ	MRYKVLFKTERTQ 940	
DB	1791	-----MEVLLASKARAEBSRS-----	TSEKSK 1813	
QY	941	AAIILQKFTFRYLAQKQLYQSVHSTIITQWRRAQOLGRQ-HRQRF	VELREAAIFLQRIW 999	
DB	1814	QRLEAAGRFRELAEEAA-----RLRALAEAKRQ--QLAEEDA	ARQRAE 1857	
QY	1000	RRRLFAKLLAAETARLQRSQKQAAASYIQMOWRTYQLGHIQRH	FEFLRQDILLMFQVR 1059	
DB	1858	AERVLAEKLAATGEATRL-----KTSBAETALKEAENRRLRLA	EDAEAFORRL-----EE 1909	
QY	1060	RMRSKWSMLEQRKEFOOLKRAAINIQQRWAKL-----SMRKNADY	LALRSSVLKVQAY 1114	



Db 2071 --KASFEAAAGKABLE---LELGRIRSNADTMRKELAEQEAARQORLAAEEERQRRRE 2125  
 QY 1056 ---FVORMRKSWMLQOK---BFOOLKRAINIQQWRRAKLSMRKCNADYTLALRSV 1108  
 Db 2126 ABEVQVRSIAEBAARQKVALEVERLK-AKVBEARKLRERARQESARQIQLAQEAQA 2184  
 QY 1109 LKQVAYRKATIOMRIDRNNHYSLRKNVTCLOQRRAIMKMRQERNYL-RLRN----- 1160  
 Db 2185 KRLQAEKA-----HAFVQQRBEELQOTLQEQNMLERLSEAEAAAR 2228  
 QY 1161 -ASILVQRYMRQOMIQDRNAYLTRKCIINVQRWRATLQ-----MRERKNYLHL 1212  
 Db 2229 AABEAAREQAREARQASQKQVEBAERLKQSAEQQAQAQAQAABAKLKEAE----- 2283  
 QY 1213 QTTTKRIQIKRAKEMKQRAEFLQKVTLLVQKRRALLQMKERQEYLHLREVTIK 1272  
 Db 2284 QEAARQAQAQALQKQAADAMEKHKFA-----BOTLKQAQVQBELTTLR---LQ 2334  
 QY 1273 LQRRPHAQSM---RFMRKRYGTQAQVSCLO-----MEWRNHL 1308  
 Db 2335 LEETDH-QKSILDEBLQRLKAETVTEAARQSQVEBELFSVRVQMEELGKIKARIEAENRA 2393  
 QY 1309 LKREKENS---FLOLRQAAITLQRRYARLNMIKQIKSYAQLKQA--ITQTRYRAKA 1363  
 Db 2394 LILRDKDQTRFLEBAEKMKQVAEEAARLSVAAQ--EAARLQALABEDLAQORALABKM 2451  
 QY 1364 MOKQVLYQ-----KQBAILKVQRRYRGNLEMRKQIEVYQKQRAVIRLQKWRSEI--- 1415  
 Db 2452 LKERMQVAQEAATLKAEBELQOQKELAQEQARRLQADKQWAQOLVEETQOFQRTLEAE 2511  
 QY 1416 --RMLCKAGVR-RIRLSSISIQKWRATVOARRQRE-----IFLSTIRKVL 1461  
 Db 2512 RQOLEMSAEARLKLMAENS-RAQARAEBDAQRFKQAEIEGKILHRTELATQEKVTL 2570  
 QY 1462 MQAFIRATL-LMRQOR-REFEMKRAAVIQR-----RFRARCMLKARDYOLIQSVI 1514  
 Db 2571 VQ-----TLETQROQSDQADRELRERAEALEREKELQKQAKLLOKS--EEMQTVQQRQI 2624  
 QY 1515 L-----VQKFRANRSMQARQEFVOLRTIAVHLQKF-----RGRKLMIEQRNCFQLLR 1564  
 Db 2625 LQETQALQKSFSEKDSLLQERFIEQB--KAKLEQLFQDEVAKAKQIQEQQ----- 2674  
 QY 1565 CSMPGFQARAGFMARKEFQALMTPEMDLI-----RQKRAKVITYRWRYG--LIR 1614  
 Db 2675 -----QRQOQMEQKQELVASMEEAARRQREAEQVRRKQEBELQRLQEQ 2718  
 QY 1615 PRQKHQGLL-----DIRKRIALQREAKA 1638  
 Db 2719 QRQOQKLLAEENQRLRERLQRLBEEHRA 2747  
 RESULT 11  
 T05200  
 myosin heavy chain F4110.130 - Arabidopsis thaliana  
 N;Alternate names: protein F4110.130  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 03-Dec-1999  
 C;Accession: T05200  
 R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hoheisel, J.; Mew  
 submitted to the Protein Sequence Database, February 1999  
 A;Reference number: Z15402  
 A;Accession: T05200  
 A;Molecule type: DNA  
 A;Residues: 1-1375 <BRV>  
 A;Cross-references: EMBL:AL035525  
 A;Experimental source: cultivar Columbia; BAC clone F4110  
 C;Genetics:  
 A;Map position: 4  
 A;Introns: 48/3; 90/3; 139/2; 191/3; 211/2; 263/3; 313/3; 359/2; 408/2; 442/2; 461/3; 49  
 A;1134/3; 1155/3; 1214/3; 1238/2; 1271/3; 1298/3; 1326/2  
 A;Note: F4110.130  
 C;Superfamily: myosin MYO2; myosin motor domain homology  
 C;Keywords: P-loop

F:63-714/Domain: myosin motor domain homology &lt;MMOT&gt;

Query Match 3.2%; Score 300.5; DB 2; Length 1375;  
 Best Local Similarity 21.5%; Pred. No. 1.2e-08;  
 Matches 263; Conservative 178; Mismatches 406; Indels 379; Gaps 65;  
 QY 320 NEMKRTSEONRYVEQOQIQIKAKDLNSSSSSEASLAGQOEFLFNHSEIL-AQSSEFNLAH 378  
 Db 201 NAKTVRNNSSRFGKVFBIQF---DTNGRISG---AAIRTYLLERSRVVRIIDPERNVH 253  
 QY 379 EYGRKSVKGS-----PVKNPHKRSHSLSFS-----DAPSNESLYRN-----ETVAISPP 423  
 Db 254 CFYQLCAGSDAEBKYLKSNP--RQPHYLNQSKTYELEGVSSAAEYKNTREANDIVGISQD 311  
 QY 424 KQORVEDTTLPRSAAPANASARSSSAHAWPRAQSKFKLAQTMSLMKPKPATPRKVRDRTSI 483  
 Db 312 BOEGT-FETLAAIHLGNVEFSSGREH-----DSSVVKDOPESRH-----HL 351  
 QY 484 OPSVKLY--DSELVMOT-CINP-----DPEAATTTIDPPLASTMYLDEQAVDR 528  
 Db 352 QMAADLFKCDANLLASICTRSILITREGIIIKALDPNAAVISRD-TAKTVY----- 402  
 QY 529 HQADPKMLNALVISPADLDADLNKKIDVGKL-----FNEVRNKE-----LVVAPTKBQS 579  
 Db 403 --AHLFDWL--VDKINKSVQDPSERFOIGVLDIYGFECFKNNSPEQFCINPANEKLOQH 458  
 QY 580 MN-VLTTKVLETLKAAVELFSEQMRPLCSKVAVYVKNQALRIRSDRNHLHDVVMQRTI 638  
 Db 459 FNEHVFQKQDEYRKBELNWSYIE-----FIDNQDV-----LDLIEKPI 498  
 QY 639 LELL--CNPLWLRLGLEVVFGKIQMOSNRDIVGLSTFILNRLFRNKCEBORYSKA 694  
 Db 459 GVIALLDKACMP-----RSTHESFSMKLF-QNFRPHRLEKFKFSET 540  
 QY 695 YTLFEYEAETIKHSLQKILFLPLDOAKQRIYKH-----NPLFV-----KKSPHKE 744  
 Db 541 DFTLSHYAGK-----ATFLDKNRDVTIVEHCNLLSSSKCPFVAGIFPSAPEES 588  
 QY 745 TKDILLRPS-----ELLANIGDITRELRGLVLOHROTFLDEDFYAFNNLAV--DL 795  
 Db 589 TRSSSYFSSVSSRFKQOLQALMETLSK--TEPHYVRCVKPNSLNR-PQKPESLSVLHQL 645  
 QY 796 RDGVRLTVRVBEILLRDDLTRQ-----LRVPAL--SRQRIENVKLALGALGEA 842  
 Db 646 RCGGVLEAV--RISLAGYPTFRNYSDFVDRGLLAPEFMDENDEQALTEKI--LSKGLG 702  
 QY 843 NFQLG-----GDIAAQDIVDGHREKTSLLMLQIYKFRSPKFAAATVLOKWRERHW 894  
 Db 703 NYQLGRTKVFLRAGQIG--ILDSRAEVLN-----ASAR----- 734  
 QY 895 LHVVIQRRIR-----HKELMRHRAATVIOAVFRGHQMRKYVKLFKTER-TQAAIILQKFT 949  
 Db 735 ---LIQRRLETFVTHQNFISARASAIQYACRGCLSR---NAYATRRNAAAAVLQKHV 788  
 QY 950 RYLAQQLQVSHSIITIQRWBAQQLGRQHRQFVELREAAIFLQRIWERELFAKLL 1009  
 Db 789 RRLWS-----RCAFKVLVSAALVLOGCIR----- 812  
 QY 1010 AAATFALORS-QKQAAAASYIQMWRTYQLGRIQRHEFLQRDLIMFVQRMRSKWSML 1068  
 Db 813 --ADSTRKFSHQEHRASLIQAHWHIKF---RSAFRHQSSIIAIQCRWQKLA-- 864  
 QY 1069 EORKEFOOLKEA-----INIQOR-----WRAKLSMRKCNADYLA----- 1103  
 Db 865 --KKEFRKLQVANEAGALRLAKTKLEKRLDLEWRILQLEKRLTSGEAKSSEISKLOK 922  
 QY 1104 -LRSSVLKQVAYRKATIQMRIDRNHYSLRKNVTCLOQRRAIMK-----WREQ 1151  
 Db 923 TLESFSLKQDAARLAIN-----ECKNAV-LEKQIDISMKEKSAVERELNGMVEL 972  
 QY 1152 RENYLRLNASILVQKRYMRQOMIQDRNAYLTRKCIINVQRWRATLOQRRE-RKNYL 1210  
 Db 973 KCONALLKNSMNSLEKKNRVLE-----KELLNAKTCNNVTLOKLEAEKRC 1019

QY 1211 HLQTTRKRIQIKFRAKREMKQRAEPLQLKKVT-----LVVQRRRALLQMRKEQEY 1263  
||| :  
Db 1020 ELQTSVSQLEELK---SHLENENOVLMQTLITSPERIGQILGEKHSSAVVPQAQDRRSV 1076  
||| :  
QY 1264 LHLREVTTKLQRFFAQSMMRWAKRYGTGAASVCLOMEWR-----N 1306  
||| :  
Db 1077 FENYELLSRC-----IKENLGNDDKPLAACVITYKCL-LHWRAFESSESTAFNIIEGIN 1130  
||| :  
QY 1307 HLLKRER-NSFLQRLQAAAITLQRRYRARLNMIKOLKSAYLAQLQAAITIQTRYRAKKAMQ 1365  
||| :  
Db 1131 EALKNLRSNFN-----ASAQRSRAAGVKSPFKLHGP-DDGASHIEARYPA----- 1179  
||| :  
QY 1366 KOVVLYQKQREAIL-KVORRYRGNLE 1390  
||| :  
Db 1180 ---LIFKQOLTACVEKIYGLIRDNLK 1202  
||| :

RESULT 12

A59254 myosin heavy chain 12, splice form 1 - human  
C:Species: Homo sapiens (man)  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Sep-2000  
C:Accession: A59254  
R:Meurers, B.  
submitted to GenBank, September 1996  
A:Description: The complete cDNA for human myosin heavy chain 12, a class V myosin.  
A:Reference number: A59254  
A:Accession: A59254  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1855 <MW>  
A:Cross-references: GB.Y07759; NID:g1770740; PIDN:CAA69035.1; PID:g1770741  
A:Experimental source: isolate 33S; dev stage adult; tissue type skin; cell type fibroblast  
C:Comment: For an alternate splice form of the same gene see PIR:B59254.  
C:Genetics:  
A:Gene: MYH12  
A:Map position: 15q21  
C:Superfamily: myosin MYO2; myosin motor domain homology  
F:72-751/Domain: myosin motor domain homology <MMO>

Query Match 3.2%; Score 300.5; DB 2; Length 1855;  
Best Local Similarity 19.2%; Pred. No. 1.9e-08;  
Matches 289; Conservative 224; Mismatches 545; Indels 450; Gaps 64;

QY 559 KLFNEVRN--KELVWAPTKEEQSMNVLTGYRLETLRKAAYVELFESEQMRLPCSKVAVYN 616  
||||| :  
Db 410 KLFNMIVNVNQALHSAYKQHSFGVLDTYGTFEFINSFEQP-----CIN--YAN 458  
||||| :  
QY 617 KOALAIRSDNLHL-----DVVMQRTILELLL-----CFNPLMLRLGLEVVFGKIQM 664  
||| :  
Db 459 E---KLQQFNMHVFVPLEQEYEMKEQIPWTLLDFDYNQPCINLTIESKLGILLDLDECKM 515  
||| :  
QY 665 QSNRDIVGLSTFILNFLRNKC---EEQRYIS-KAY-----TILT 698  
||| :  
Db 516 PKGTDDETWAQKLYNTLH--NKCALFEKPRLSNKAFTIHFADKVEYOCEGFLEKNKDVTFF 573  
||| :  
QY 699 BEYAETIKKHSQKILFLPFLDQAKQRI-----VKHNPCLFVKKSP----- 741  
||| :  
Db 574 EEQIKVKSXKPK----MULPELDQDEKASPTSATSGSRTPLTETPAKTGRPGQMAK 629  
||| :  
QY 742 -HKETKDIIILRPSELLANIGDITRELRLRGYLQHROQTFLEDFDYAFNNLAVDLRDGV 800  
||| :  
Db 630 EHKTGVGHQFRNSLHL-----METLNATPHYVRCIKPNDFKL-PFTFDEKRAVO 679  
||| :  
QY 801 LFRVVEVILLRDDTLRQLRPA-----ISRLQIRFNVKIALG----- 837  
||| :  
Db 680 QLRAQCVL-----ETIRISAAGFPSEWTYQBEFSRVRLMKQKDVLSDRKQTKCNVLE 732  
||| :  
QY 838 --ALGANPQLGGDIAAQDVGHRKETSLLMQLIY--KPRSPKFHAATAVTVLQKWRRH 893  
||| :  
Db 733 KLILDKDKYQFG-----KTIFFRAGQVAVLEKLRADKLRAACIRIQK-TIRG 779  
||| :









*This Page Blank (uspic,*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:36:33 ; Search time 23 Seconds  
(without alignments)  
4213.155 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVSPVLEVACKETQLLI.....FISSVYAFDTILCKLQIDWF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	324	3.4	1407	1 TRHY_RABIT
2	323.5	3.4	1549	1 TRHY_SHEEP
3	317.5	3.3	1829	1 MYSA_CHICK
4	303.5	3.2	2779	1 LVA_DROME
5	301.5	3.2	1855	1 MYSA_HUMAN
6	301	3.2	4687	1 PLE1_RAT
7	295.5	3.1	1828	1 MYSA_RAT
8	294	3.1	4684	1 PLE1_HUMAN
9	287	3.0	4473	1 PLE1_CRIGR
10	284.5	3.0	1805	1 HWM2_MYCGE
11	281	3.0	1742	1 MY5C_HUMAN
12	277	2.9	1818	1 HWM2_MYCPN
13	277	2.9	1898	1 TRHY_HUMAN
14	274	2.9	1498	1 GOA3_HUMAN
15	271.5	2.9	1853	1 MY5A_MOUSE
16	270	2.8	2442	1 CEP2_HUMAN
17	267.5	2.8	2663	1 CENE_HUMAN
18	260	2.7	1979	1 TRIA_HUMAN
19	260	2.7	2245	1 MY5J_DICDI
20	258.5	2.7	2238	1 GOA4_MOUSE
21	257.5	2.7	1849	1 MY5B_HUMAN
22	254.5	2.7	1411	1 EEA1_HUMAN
23	254	2.7	3210	1 CENF_HUMAN
24	246.5	2.6	2022	1 ANTI_ONCVO
25	244	2.6	1846	1 MY5B_RAT
26	243.5	2.6	1453	1 Y373_BOVIN
27	238.5	2.5	3674	1 SPOR_HUMAN
28	234.5	2.5	1941	1 MYH2_HUMAN
29	232.5	2.4	1679	1 GCC2_MOUSE
30	230.5	2.4	1211	1 SRCC_PSEAE
31	229	2.4	1539	1 Y373_HUMAN
32	228.5	2.4	1679	1 Y109_YEAST
33	228	2.4	2611	1 BPLE_MOUSE
				Q91zu8 mus musculus

#### ALIGNMENTS

```

RESULT 1
TRHY_RABIT
ID TRHY_RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trichobyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC layer of the epidermis. It later becomes cross-linked to KIF by
CC isodipeptide bonds. It may serve as scaffold protein, together
CC with involucrin, in the organization of the cell envelope or even
CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -!- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -!- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely
CC alpha-helical, configured as a series of peptide repeats of
CC varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIF directly by ionic interactions.
CC Domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC different species.
CC -!- PTM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -!- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; Z19092; CAA79519.1; -.

```

DR PIR: S28589; S28589.  
 DR HSP: P02633; 4ICB.  
 DR InterPro: IPR001751; CabP\_S100.  
 DR Pfam: PF00036; EF-hand; 1.  
 DR ProDom: PD003407; CabP\_S100; 1.  
 DR ProSITE: PS00018; EF HAND; 1.  
 DR ProSITE: PS00303; S100 CABP; 1.  
 KW Keratinization; Calcium-binding; Repeat; Citrullination.  
 FT DOMAIN 1 91 S-100 LIKE.  
 FT CA BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 FT CA BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;  
 Query Match 3.4%; Score 324; DB 1; Length 1407;  
 Best Local Similarity 21.5%; Pred. No. 2.7e-10;  
 Matches 263; Conservative 211; Mismatches 455; Indels 294; Gaps 47;  
 QY 680 RLPRNKCEORYSK-----AVLTTEYAEATIKKSLQKILFLPDLQAKQKRV 729  
 Db 173 RLQORQLEERAESEQLRRKGDABEFIE--EQLRRQEQELKRELREEQORRE 230  
 QY 730 KHNPCFLVKSPKHKTDILL-----RFSSELANIGDITRELRLGYVLQHRQTF 780  
 Db 231 QHERAL-----QEEQELLRQRRWREPQEQQLRRELEIRERQORLEQERREQOL 283  
 QY 781 LDFDYAFNNLAVDLRGVLTTRVVEVI-----LLRDDLTQLRVPAISLQLRIFNVKLA 835  
 Db 284 RRE-----QRLEQERREQOLRRELEIRERQORLEQERREQORLEQERREQOL 336  
 QY 836 LGALGRANFOLGPD-----TAAODIVDGHREK--TSLILWOLIVK-----FRSPK 878  
 Db 337 LEIIREREQLEQERREQLLAEEVREQARBERGESUTRRWQOLSEAGARQSKVYSRPR 396  
 QY 879 FHAATVLQKWRHHVHVVIQRIHKLMLRHRRAATVTOAVFRGHOMKRV----- 931  
 Db 397 RQEQSLRQDERR-----QRQERRELEQARQOQWCAEESERRRQLSARPSLR 449  
 QY 932 --KLFKTERTCQAIILOKFRVYLAQOLQSHYSIITIQWRAQOL-----GRQH 981  
 Db 450 ERQLRAEERQEQORFEEFEEFQRRERQEQOLFLSEELQORRERAAQOLBEDSFQEDR 509  
 QY 982 RQRFVLERAAIF--LQRIWRR--LFAK-----LKLAAAT 1014  
 Db 510 RRRQEQORPGOTWQVQLQEAQRRTLYAKPGQEQQLREELQREKRRQERREYREB 569  
 QY 1015 ARLQSQKQAAASYIQWRTYOLGRIOHREHFRDRLIMFVQ----- 1058  
 Db 570 EKLQREDEKRRRQERERQYR--ELEELRQEQQLDRKLREELQORLEERLRRQERE 627  
 QY 1059 RRMESKWSMLEQ-----RKEFQOLKAAINIQQ----- 1086  
 Db 628 RKLREELQLEQEQELQERERKLREELQLEERQELQERERKLREELQLEEREE 687  
 QY 1087 RWRAKLGMKCNADYLAIRSSVLQVQYRKATIQ-----MRIDRNHYISLKN 1134  
 Db 688 RLRRQERARKLREELQLEQEQELQERERKLREELQLEERQLEERLRR--QERERKL 745  
 QY 1135 VICLQ-----QLRAIMKMRQENYLR-----LRNASILVQKYRMQMIQDRNAYLR 1184  
 Db 746 EQLQLESEERLR-----RQREQLRRERDRKFRREELQLEERERLRR--QERERKL 799  
 QY 1185 TRKCIIVQWRATLQWRERKKNYLHOTTTKRIQIKF--RAKEMKQRAEFLQLKKVT 1243  
 Db 800 EEEQLQLEERERLRQ--ERERKLREELQLEERERLRRQERERKLREELQLEEREE 858  
 QY 1244 LVVQKRRALIQ-----MKERQEVYHLREVITKLQRFRHAQKSRFMFAKYRGTAQAVSC 1299  
 Db 859 ELRQERARKLREELQLEQEQELQERDRKLREELQLEERQLEERQLEERERQOL 918  
 QY 1300 LQWHRNHLKRRENSPLQROAATLQ-----RYYRARNMIKOLKSYAOLKQ--AAIT 1353

Db 919 LQ--ESEERLRRQERERKLREELQLEERERLRRAR-----KLREELQLEEREE 972  
 QY 1354 IOTRVRAKAMOKQVLYQKOREAIIKQVRRYRGNLEMRKQIEVYQKORQAVIRLQKWR 1413  
 Db 973 LRQERARKLREELQLEEREEQL-----RQERDRKFRREELQLEERERLRRQERDR 1026  
 QY 1414 SIRDMLCKAGYRIRLSSLSIQ-----RKWRATVQARRQRIHFLSTIRKVLMAQAFIR 1467  
 Db 1027 KFEEREE-----QLRRQLELEQFRQERDRKFRLEQIRQEK--EKQLRQERDRKFR 1078  
 QY 1468 ATLLMQORREFEMKR-----RAAVVIQRRFRARCAMLKARQYQLI--QS 1511  
 Db 1079 EQQRQREQLRRERDRKFRREELQLEERERLRRQERAR-----KLREELQLEER 1134  
 QY 1512 SVILQKFRANRSMQARQEFVQLRTIAVHLQKFRGKFLMIEQRCNFOLLRCSPMGFQ 1571  
 Db 1135 EQLLRQERDRKFRREELQLESEERLRQERERKLREELQLEERERLRR-----Q 1188  
 QY 1572 ARARGFMARKRFOALMTPEMDLIRQKRAKVIRYWRGVLIRRRQKHQGLDIRKRIQ 1631  
 Db 1189 ERAR--KLREELQLEEREE--LRQERARKL-----RREELQLEERERLRRQER 1229  
 QY 1632 LRQF-----AKAVNSVRCKVQEAVERFLRGHFIASDALAVLSQDLRLSRTVPHLLMWCSEF 1686  
 Db 1230 LRQERDRKFRREELQLEERERDRKFRREELQLEERERLRRQER----- 1279  
 QY 1687 MSTFCYCIQAIRSEVVDKOLIE 1709  
 Db 1280 -----RARKLREELQLEF 1294  
 ID TRHY SHEEP STANDARD; PRT; 1549 AA.  
 AC P22793;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Trichohyalin.  
 GN TH.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 CX NCBI TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93260019; PubMed=7684041;  
 RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;  
 RT "Analysis of the sheep trichohyalin gene: potential structural and  
 RT calcium-binding roles of trichohyalin in the hair follicle.";  
 RL J. Cell Biol. 121:855-865(1993).  
 RN [2]  
 RP SEQUENCE OF 1016-1549 FROM N.A.  
 RC STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=Wool follicles;  
 RX MEDLINE=90130632; PubMed=2298812;  
 RA Fietz M.J., Presland R.B., Rogers G.E.;  
 RT "The cDNA-deduced amino acid sequence for trichohyalin, a  
 RT differentiation marker in the hair follicle, contains a 23 amino acid  
 RT repeat.";  
 RL J. Cell Biol. 110:427-436(1990).  
 CC -!- FUNCTION: Intermediate filament-associated protein that associates  
 CC in regular arrays with keratin intermediate filaments (KIF) of the  
 CC inner root sheath cells of the hair follicle and the granular  
 CC layer of the epidermis. It later becomes cross-linked to KIF by  
 CC isodipeptide bonds. It may serve as scaffold protein, together  
 CC with involucrin, in the organization of the cell envelope or even  
 CC anchor the cell envelope to the KIF network. It may be involved in  
 CC its own calcium-dependent postsynthetic processing during terminal  
 CC differentiation.  
 CC -!- SUBUNIT: Homodimer (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;

CC CC Comment=Additional isoforms seem to exist;  
 CC CC Names=Long;  
 CC CC IsoId=P22793-1; Sequence=Displayed;  
 CC CC Names=Short;  
 CC CC IsoId=P22793-2; Sequence=VSP\_000847, VSP\_000848;  
 CC CC TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as  
 CC CC the inner root sheath (IRS) of hair follicles and medulla, and in  
 CC CC the epithelia of the tongue, hoof and rumen.  
 CC CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand  
 CC CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely  
 CC CC alpha-helical, configured as a series of peptide repeats of  
 CC CC varying regularity, and are thought to form a single-stranded  
 CC CC alpha-helical rod stabilized by ionic interactions. Domain 6 is  
 CC CC the most regular and may bind KIF directly by ionic interactions.  
 CC CC Domains 5 and 7 are less well organized and may induce folds in  
 CC CC the molecule. Domain 9 contains the C-terminus, conserved among  
 CC CC different species.  
 CC CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are  
 CC CC probably converted to citrullines by peptidylarginine deiminase.  
 CC CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100  
 CC CC family.  
 CC CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC -----  
 CC CC EMBL; Z18361; CAA79165.1; -;  
 CC CC EMBL; X51695; CAA35992.1; -;  
 CC CC PIR; A40691; A40691.  
 CC CC HSP; P02633; IIG5.  
 CC CC InterPro; IPR001751; CaBP S100.  
 CC CC InterPro; IPR002048; EF-hand.  
 CC CC Pfam; PF00036; EF-hand; 1.  
 CC CC Pfam; PF01023; S\_100; 1.  
 CC CC ProDom; PD003407; CaBP S100; 1.  
 CC CC PROSITE; PS000018; EF HAND; 1.  
 CC CC PROSITE; PS00303; S100 CBP; FALSE NEG.  
 CC CC Keratinization; Calcium-binding; Repeat; Citrullination;  
 CC CC Alternative splicing.  
 CC CC DOMAIN 1 91 S-100 LIKE  
 CC CC CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 CC CC CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 CC CC DOMAIN 413 832 14 X 28 AA APPROXIMATE TANDEM REPEATS.  
 CC CC REPEAT 413 448 1-1.  
 CC CC REPEAT 449 476 1-2.  
 CC CC REPEAT 477 504 1-3.  
 CC CC REPEAT 505 532 1-4.  
 CC CC REPEAT 533 560 1-5.  
 CC CC REPEAT 561 588 1-6.  
 CC CC REPEAT 589 616 1-7.  
 CC CC REPEAT 617 644 1-8.  
 CC CC REPEAT 645 678 1-9.  
 CC CC REPEAT 679 706 1-10.  
 CC CC REPEAT 707 742 1-11.  
 CC CC REPEAT 743 771 1-12.  
 CC CC REPEAT 772 796 1-13.  
 CC CC REPEAT 797 832 1-14.  
 CC CC DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.  
 CC CC REPEAT 938 961 2-1.  
 CC CC REPEAT 962 985 2-2.  
 CC CC REPEAT 986 1021 2-3.  
 CC CC REPEAT 1022 1044 2-4.  
 CC CC REPEAT 1045 1067 2-5.  
 CC CC REPEAT 1068 1090 2-6.  
 CC CC REPEAT 1091 1121 2-7.  
 CC CC REPEAT 1122 1144 2-8.  
 CC CC REPEAT 1145 1167 2-9.  
 CC CC REPEAT 1168 1197 2-10.

FT REPEAT 1198 1227  
 FT REPEAT 1228 1250  
 FT REPEAT 1251 1273  
 FT REPEAT 1274 1296  
 FT REPEAT 1297 1319  
 FT REPEAT 1320 1342  
 FT REPEAT 1343 1368  
 FT REPEAT 1369 1391  
 FT REPEAT 1392 1416  
 FT REPEAT 1417 1439  
 FT REPEAT 1440 1461  
 FT REPEAT 1462 1484  
 FT REPEAT 1485 1507  
 FT VARSPLIC 1145 1197  
 FT VARSPLIC 1251 1273  
 FT CONFLICT 1399 1399  
 FT SEQUENCE 1549 AA; 201173 MM; E72PB9PF1326E54E CRC64;  
 SQ  
 Query Match 3.4%; Score 323.5; DB 1; Length 1549;  
 Best Local Similarity 19.3%; Pred. No. 3.3e-10;  
 Matches 280; Conservative 249; Mismatches 504; Indels 421; Gaps 60;  
 QY 282 RDKLVGSPLRKYSMSK-DLSLLSPQTKYALQGMPLNE-----M 322  
 DB 308 ROKVYSPHPHQEQRRQEQELLERQEQQISEVQSLQEDQGRQRLKQRYDQNRW 367  
 QY 323 KIRTEQNRY-----QEQQIQIKAKDLNSSSS-----SPASLAGOQEF 361  
 DB 368 QLEERSQRRYTVAKPAQRQVREEQRLKEKLQREKRQERQYREVELQREER 427  
 QY 362 LFNHSEILAQSFRNLHEVGRKSVKGVKVPKPHKRSHE-----LSFSDASNSLYNE 416  
 DB 428 LQREEQLOREER-----EKRRQEREKQYLEKVELWEELQOREE 468  
 QY 417 TVAISPPK-KORVEDTTLPRGAAPANASARSSAHAWPHAQSKFKLAQTMSLMKKPATP 475  
 DB 469 REKRQEREKQYLEKVEL-REEEQLOREK-----RRQEREKQYLEKVELQEEBQLO 521  
 QY 476 RKVRTSTQPSVKLYDSLYMOTCINPDPFAATTIDPFLASTWYLDQAVDRHQADPKK 535  
 DB 522 REEREKRQEREKQYLEKVELQ-----BEEQLOREKRR 557  
 QY 536 WLNALVSPADLDADLNKIDV--GKLFNVRNKLVAFTKEBQSNLYLTXYRL--E 589  
 DB 558 -----QEREKQYLEKVELQEEQLOREKRR-----RQEREKQYLEKVELQEE 602  
 QY 590 TURKAAVELFFSBQWRLPCSVAVTVNKOALRIISDR-----NLHLDVVMORTILEL-- 641  
 DB 603 QLOREKRRQEREKQYLEKVELQEEBQVQREKRRQEREKQYLEKELQORELQOE 662  
 QY 642 ---LLCFNPLWLRLGLEVVFGKIQOMQSNRDLVGLSTFILNRLFNKCEBQRYSKAYTILT 698  
 DB 663 EQALLREKRRQEREKQYLEKVELQEEQ-----LQEREKRRQERE-ERQYLEK 713  
 QY 699 BEYAEITIKKHSLOKILFLPLDQAKOKRIVKHNPCLFVKKSPHKBTXDI-----LLRFS 753  
 DB 714 EELQ--QEERLOREKQLOREDREKRRQVREK---YLEEELQEEEDRLQEKQLLEED 768  
 QY 754 S-----ELLANIGITITRELRLGYVLQHRQTFDLDFDYAFNNLAVDLRDGVRLTRV 804  
 DB 769 REKRQYLEKVELQEEQLOREKRRQEREKQYREEL-----LREERLHRK 815  
 QY 805 VEVI-----LLRDDLTQRLQVPAISELQIFNVKLALGALGEANF---OLGDDIAAQ 853  
 DB 816 EQLORECEKRRQELERQLEEEBELQDLR-----KQFRDDQHQNEVRNS 863  
 QY 854 DIVDGHRE-----KTLSSLWQLIYKFRSPKFAAATVLQ-----KWR 892  
 DB 864 RYVSKHREKESRQLDDSW-----VRESQFQDRLRPLQDEKREKREKQESRQKEDS 918  
 QY 893 HWLVVIOQRIRHKELMRHRAATVIAVFRGHQMRK--YV-----KLFKTERT 939

```

Db 919 QPFAQLLEEQKTERDRKFRREEQQLLKGQREKIRYLEEDRKFRREEQQLRLRRE 978
QY 940 Q-----AAIILKFTYRYLAQQLYQSYHSIIITIQWRAQQLGQRQR--FVE 987
Db 979 QQLRQERDRKPREELSRQERDRKPREEQQLLQ-----REQLRQERDRKPRE 1027
QY 988 ----LREAIFLQRIWRRLF-----AKLLAAATLRLQSQVQQAASVYIQMW 1034
Db 1028 BEQLLQEREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1087
QY 1035 RYVQLGRIQRHFLQRIWRRLF-----RMRKSWMLQERKFEQQLKRAAIN 1083
Db 1088 RQERNKPREEQQLLREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1145
QY 1084 IQQRWRAK-----SMKKNAD-----YLAIRSVLVQAVRKAQIQRIDRNHY 1129
Db 1146 KXFEBAQILKREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1201
QY 1130 SLRKNVICLOQLRAIMKREORENVRLRNASILVQRYMRQOMIQDRNAVLRTRKCI 1189
Db 1202 REEQQLQERKL-----RQEREPQLQER-----DRKFBEBEQQLRQERDRKPREEQQLRLREQQL 1249
QY 1190 INVQRWRAQLQRRERKNYLHLOTTTKRIQIKFRAKEMKORABFLQKLVVQKR 1249
Db 1250 --RDRKFEBAQILKREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1296
QY 1250 REALL--QMKREBYLHREVTIKLQRRHAQ-----KSMRFRKRYRGTQAAVSLQ 1301
Db 1297 DRKFEBAQILKREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1345
QY 1302 MEWRNHLKRRNSFLQRAAIIQRRYRRLNMIKQLSYAQKAAITITRYRAK 1361
Db 1346 FHEKHLRLREEQ--QLRQERLE-----GVFSQEQQLRREBQEQERQQRDR 1393
QY 1362 KAMQKQVLYQKQRAIIVQRYRGNLEMKQIEVYQKQAVIRIQKWRSTDRMLC 1421
Db 1394 KFLREBQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1440
QY 1422 KAGYRIRLSSLSIQKRWATVQARRQREIFLSTIRKVLQAFIRATLLMRQRRRPFEM 1481
Db 1441 -----LQQRYAEQFAFE-----EKRRQERQLRQERDRKPREEQQLRLREQQL 1476
QY 1482 KRAAVTVQRRFRCAMKARQYQLIQSSVLVQKFRANRSMKQARQBFVLRTTAV 1541
Db 1477 -----RKPREEQQLRQERDRKPREEQQLLQEREBQQLRQERDRKPREEQQLRLREQQL 1507
QY 1542 HLQKFRGKRLMIE 1555
Db 1508 --VWEEDKRRQVIE 1520

```

## RESULT 3

```

MY5A CHICK
ID MY5A CHICK STANDARD; PRT; 1829 AA.
AC Q02440.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle) (Myosin
  heavy chain p190) (Myosin-V).
GN MYO5A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93012002; PubMed=1383040;
RA Sanders G., Lichte B., Meyer H.E., Kilimann M.W.;
RT "cDNA encoding the chicken ortholog of the mouse dilute gene product.

```

```

RT RT
RL FEBS Lett. 311:295-298(1992).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1154-1163.
RC TISSUE=Brain;
RX MEDLINE=93107155; PubMed=1469047;
RA Repreafico E.M., Cheney R.E., Matteoli M., Nascimento A.A.,
  de Camilli P.V., Larson R.E., Mooseker M.S.;
RT "Primary structure and cellular localization of chicken brain
  myosin-V (p190), an unconventional myosin with calmodulin light
  chains."
RL J. Cell Biol. 119:1541-1557(1992).
CC -!- FUNCTION: Processive actin-based motor that can move in large
  steps approximating the 36-nm pseudo-repeat of the actin filament.
  May be involved in melanosome transport, or alternatively, it may
  be required for some polarization process involved in dendrite
  formation (By similarity).
CC -!- SUBUNIT: May be a homodimer, which associates with multiple
  calmodulin or myosin light chains.
CC -!- SUBCELLULAR LOCATION: Golgi-derived cytoplasmic membranes
  (potential).
CC -!- TISSUE SPECIFICITY: Neuronal and nonneuronal cells of the brain.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 dilute domain.
CC -!- SIMILARITY: Contains 6 IQ domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
EMBL; X67251; CAA47673.1; -.
DR EMBL; Z11718; CAA77782.1; -.
DR PIR; S19188; S19188.
DR HSP; P10587; 1BR2.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSTINHEAVY.
DR ProDom; PD003376; DIL; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 6.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 6.
DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
  Coiled coil; Phosphorylation.
KW MYOSIN HEAD-LIKE.
FT DOMAIN 1 766
FT DOMAIN 767 789
FT DOMAIN 790 814
FT DOMAIN 815 837
FT DOMAIN 838 862
FT DOMAIN 863 887
FT DOMAIN 888 915
FT DOMAIN 915 1239
FT DOMAIN 1239 1419
FT DOMAIN 1419 1661
FT DOMAIN 1661 1700
FT NP BIND 163 170
FT DOMAIN 644 666
FT MOD RES 1734 1734
FT CONFLICT 1142 1142
FT CONFLICT 1142 1142
SQ SEQUENCE 1829 AA; 212381 MW; 0538B278DFC09F6E CRC64;

```

Query Match 3.3%; Score 317.5; DB 1; Length 1829;  
 Best Local Similarity 18.3%; Pred. No. 8.9e-10;  
 Matches 351; Conservative 271; Mismatches 670; Indels 623; Gaps 82;







Db 621 RPPASGDAQEQKQVHPGSHVSELTOTETEEDSS-----GETLSQRLERLEL 671  
QY 773 VLQHQTEDEFD-YAFNNLAVDLR-----DGVRLTRVVEVILLRDLTLRQLRPAISR 825  
Db 672 FTQGEVLDKLEQSAENLQARLEBSSSLQLQREKDLISSTSTSNLQELSS 731  
QY 826 LQRI FNVKLALGALGANFQAGDIAAQDIVDGHREKTLISLWQIYKF-----RSPKEH 880  
Db 732 MQRSEVATLDA-GE-----GGPVLFEK-----EKSLSKLSLELAEYKANDROAKEN 780  
QY 881 AAATVLOKWRH-----WLHVIOIRIRHKEIMRRHRAATVIOAVF-----RGHOMRY 930  
Db 781 VSKLAKAKNCHTQSLSELHKV-----KEASTAVETVIVETVAVTAPNGKALARY 833  
QY 931 VKLFXTERTAAIILQKFTTERRYLAKOLYQSYHSIITI-----QRWRAQQLGRHQRP 985  
Db 834 BOL-NAQNAELKANVIRLROELDELESPEPEAPLAIVGSOSQREDEILQLOSO-----887  
QY 986 VELREAAIFLQIRWRRILFAK-----KLAAAEATARIQSQKQAAASYIQMOWRTY-QLG 1040  
Db 888 ---LEDARSLAQEQOQIBEQVDQIKELRQTEAQQLVARQSAITQLQSEQFDQLL 944  
QY 1041 RIQHEFLRQDLINFLVORMRKSMLEQR-----KEFOQL-----KRAAINIQ 1085  
Db 945 NSKEMSEKOLEQOQTRIRRELEAREASLEGELSILQTLVAEQKQQLIESVSEHALNLK 1004  
QY 1086 --QRWRKSLMRKCA-----DYL--ALR--SSVLKVCARVKATIOWRIDENHYYSRKVNI 1136  
Db 1005 MLEQLAQEBELRBAKEDPDOLRRLRSLVSLVQOQVRELT-----SSQETVD 1053  
QY 1137 CLOQLRALI-----MKMREORENLYRLNASILVKRYRMR---QOMIDRNAY 1182  
Db 1054 ALNQIOEQYQGLEHAHKEEQFKNELRB---KLKVALNLKKTQDNADLQKVQELISQ 1110  
QY 1183 LRTRKCI-----NVQRWRATLOMERKKNYLHLQTTKRIQIKFRKR 1227  
Db 1111 LQEQQLVQKEEVEREPIDVNRHVEQLQOQVSKLNEDLKAKIHLNLR-----DALR 1164  
QY 1228 EMKQORAFLOLKKVTLVQKRRAL-----LQMKER-----QEVYHLREV 1269  
Db 1165 QKQIOEQEQL-----IQERDAELOANLVSKELRERQEQVQFOLQGENSLREE 1218  
QY 1270 TIKLQRRPH-----AQKSMRFMEAK--YRGTQAAVSLQW-----1302  
Db 1219 ISKLEETHNLQGVNEBPTAVEDLRLQLEAKSKFKFKSLIKLRNATIQSLQRELOQL 1278  
QY 1303 -----HWRN-----HLRKERNFLQLOAAITLQRYRA-----RLNMIKOLKS 1343  
Db 1279 QOQDQSEVHVNRANAHEQLRLKEDAEITALRQELIKLERSRAAGEGDDTITKTHQLL 1338  
QY 1344 YAQLKQAAITITQRYRAKAMQKVLYQKQREALIKVORYRGNLEMRKQIEVYQKQ 1403  
Db 1339 ESQSQQASLOVAERELQURVOLTAQEQHALLAQOYASDKANFEN-----1386  
QY 1404 AVIRLOKWRMSIRDMRLCKAGY-----RRRLSSLSIQKRWATVQARRQREIFLSTIR 1457  
Db 1387 TIARLETHGEOAKIQEDASYIESLEAQNTLOARSALAEQASQANQO-----AASQD 1442  
QY 1458 KVRMLQAFIRATLIMQORREFFEMKRAAVIQRFF-----1493  
Db 1443 KVQILQEQ-----QKQEQEQEQKQOQOQQLQERFVBLQEQEQSQLELLTSEABES 1496  
QY 1494 RARCAMLKARQDYOLQSSVILVORKFRANSMKQARQEFVQLR-----TIAVHLQKF 1547  
Db 1497 RQIAGLRTYESLLAKHSQLTATAQAREQMSHQSQELAEALRQLQDLVKEADLHRQV 1556  
QY 1548 RQKRLMIEQRCNFOLLRCMFGQARARG-----FMARKRFOALMTPEMDLIRQKAAK 1602  
Db 1557 YDAKLAKATFELDE-LECDLNSHVERRAAETREICQOLERSQELVAQRTBELQRLNEBFQ 1615  
QY 1603 VIQRYWGYLIRR-----ROKH-----QGLIDIRKRIAQLQREKAVNSVRCKV-----Q 1647

Db 1616 EVERE-RSTLSREVTILRLQHDSEAQDVLEQLQELRMAQMDKTEMNLTQTIDALCANHS 1674  
QY 1648 EAVRFLRGFIASDALVLSQDLRLSRVTPHLLMCSEFMTFCYGINMAQAIRSEYDKOL 1707  
Db 1675 QELOALQORIAELDTIGQNTDDQV-----YIETE-NKRL 1708  
QY 1708 IERCSTRIILNAR 1720  
Db 1709 ABQLSELQQLAR 1721  
RESULT 5  
MYSA HUMAN  
ID MYSA HUMAN STANDARD; PRT; 1855 AA.  
AC Q9Y4I1; O60653; Q07902; Q16249; Q9UE31; Q9UE31;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)  
DE (Myosin heavy chain 12) (Myosin).  
GN MYOSA OR MYH12.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Skin;  
RA Meurers B.H.; Zimmermann R.; Vosberg H.P.;  
RT "The complete cDNA for human myosin heavy chain 12, a class V  
RT myosin.";  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.  
RX MEDLINE=97351514; PubMed=9207796;  
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,  
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;  
RT "Griscelli disease maps to chromosome 15q21 and is associated with  
RT mutations in the myosin-Va gene.";  
RL Nat. Genet. 16:289-292(1997).  
RN [3]  
RP ERRATUM.  
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,  
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;  
RL Nat. Genet. 23:373-373(1999).  
RN [4]  
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).  
RC TISSUE=Fetal brain;  
RA Engle L.J., Kennett R.H.;  
RT "Cloning, analysis, and chromosomal localization of myosin (MYH12),  
RT the human homologue to the mouse dilute gene.";  
RL Genomics 19:407-416(1994).  
RN [5]  
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).  
RC TISSUE=Brain;  
RA MEDLINE=95136715; PubMed=7835087;  
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,  
RA Jenkins N.A.;  
RT "Cloning and regional assignment of the human myosin heavy chain 12  
RT (MYH12) gene to chromosome band 15q21.";  
RL Cytogenet. Cell Genet. 69:53-58(1995).  
RN [6]  
RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).  
RA Edgar A.J., Bennett J.P.;  
RT "Inhibition of dendrite formation in melanocytes transiently  
RT transfected with antisense DNA to myosin V.";  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP FUNCTION.  
RX MEDLINE=99376094; PubMed=10448864;  
RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,  
RA Cheney R.E.;



```

Db      886 M---HAIYIQQCFRRMAKRELUKUKIBAR-----SVERY 918
Qy      1046 EFLR--ORDLIMFVQRMSKWSMLERKEFOOLKRAAINIQORWRAKLMSRKNADYLA 1103
Db      919 KKLHIGWENKIMQORQVD-----BQNDKYCKLVEKLNLEGIYSETEKRLSDLERLQ 972
Qy      1104 LRSVLKVOYAKATQMRID--RNHYISLRKQVVICLOQLRALIMQORQRENYLRLNA 1161
Db      973 LSEBEAKVATGRVLSQEEIAKLKDLQLEQTRSEKKCIEEHADRYKQETQLVSNLKEEN- 1031
Qy      1162 SILVQKRYMRQOMIDENAYLSTRKCIINVRWRWATLQMRERKNYHLHLOTTTKRIOI 1221
Db      1032 TLXQKEALNHRIVQAKEMTEMEKKL-VETKQLELDLNDERLYQNLLNEFSRLSE 1090
Qy      1222 KFR-----AKEMKQKQARFQLQKVT- 1243
Db      1091 RYDDLKEEMTLVHVPKPGHKRTDSTHSSNESEYIFSSIAEMEDIPSRTEEPSEKKVPL 1150
Qy      1244 ----LVVQKERRALLQMRKEQBYLHLREVITIKLQRRFHAQKSMFRMAKYRGTOAAYS 1298
Db      1151 DMSIFLKIQKRVTELEQEKQVMDQLDRKEQVL-----RSKAKEERPPQIRGALEVE 1204
Qy      1299 CLOQWHRNHL--LRKERNSEFLQRLQAAITLQRR-----YRARLANMIKOLKSY 1344
Db      1205 SLK---RQELSESEKKLKNELNELRKA---LSEKSAPEVTPGAPAYRVLMELQTSVSEE 1258
Qy      1345 AQLKQAAITQITRYAKAMQKVVL-----YQKOR-----BAIKV 1381
Db      1259 LDVRKEEVLV---LRSQVSKQEAIPKDKNTMTDSTILLEVQVKQKQGETAQAYIGL 1315
Qy      1382 ORRVGNL---EMRKQIE--VYQKORQA---VIRLOKWERSIRD-----MRLC 1421
Db      1316 KETNRSSALDYHELNEDEGLWLVYEGLKQANRLLESQLOSKSHENAEALRGETQSLK 1375
Qy      1422 KAGYRIRLSLSIQKRWATVQARRQREIFLSTIRKVRIMQAFIRATILMRQORREFEM 1481
Db      1376 EENNROOQLLAQNLQLPPEARIEASLQHEITRLTNENMLDMEQ-----LEKQDKTVRKL 1429
Qy      1482 KRRAAVIQRRFRARCAMLKARQDYQLIQSSV--ILVQKFRANRSKQARQEFVOLRTI 1539
Db      1430 KQQLKVFPAKKIGLEVQGMENISFGQIIDPIRPVNIPIRKEKDFQGMLEYKKEDEOKLVK 1489
Qy      1540 AVHLQQRFRGRKRLMEQRNCFOLLRCSPMGFOARARGFMAR-----KRFQALMTPEM 1591
Db      1490 NLILKLPKRGAVNL-----IPGLPAVILPWCVRHADYLDNDQKVRESLITSTI 1537
Qy      1592 MDLIRQKRAKVOR-----YRQGYLIR-----RQKQOGLL 1623
Db      1538 NSI-----KKVLKRGDDFETVSWLNTCRFLHCLKQYSGEGFMKHNTRQNEHCLT 1591
Qy      1624 DIRKRIQLRQEKAKVNSVRCKV-QEAVRFLRG---RPIASDAL-----1663
Db      1592 NF--DLAETQ---VLSDLAIQIYQQLVRLVLENILQPMIVSGMLEHETIQGVSQVPTGL 1646
Qy      1664 --AVLSQDLRLSTVPHLLMWCSEFMSTFCYGINAQARISEVDKQLIERCSRILN----1717
Db      1647 RKTSTSIADGETVTLDSILQNLNSFHSVMQHGMDPELIRQVVKQMPFYIIGALTANLLL 1706
Qy      1718 -----LARYNSTTV-----NTQEGG-----LVTTAOML--LRWCDDKDEI 1751
Db      1707 RDMCWSKSGMQRIRYNVSQLEWLRDNLMNSGAKETLEPLIQAAQLLVKKKTTDDDAEA 1766
Qy      1752 FNTLCTILI 1759
Db      1767 ICSWCNAL 1774

```

RESULT 6  
 ID PLE1 RAT  
 AC P30427; O08879; O08880; O08881;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Plectin 1 (PLTN) (PCN)  
 GN PLE1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX TISSUE=Glial tumor;  
 RC MEDLINE=91268156; PubMed=2050743;  
 RA Wiche G., Becker B., Lubert K., Weitzer G., Castanon M.J.,  
 RA Hauptmann R., Stratowa C., Stewart M.;  
 RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide  
 chain with a three-domain structure based on a central alpha-helical  
 coiled coil.";  
 RT J. Cell Biol. 114:83-99 (1991).  
 RL [2]  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4), AND TISSUE  
 SPECIFICITY.  
 RX TISSUE=Glial tumor;  
 RC MEDLINE=97321050; PubMed=9177781;  
 RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,  
 RA Wiche G.;  
 RT "Plectin transcript diversity: identification and tissue distribution  
 of variants with distinct first coding exons and rodless isoforms.";  
 RL Genomics 42:115-125 (1997).  
 CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and  
 microfilaments and anchors intermediate filaments to desmosomes or  
 hemidesmosomes. May be involved not only in the crosslinking and  
 stabilization of cytoskeletal intermediate filaments network, but  
 also in the regulation of their dynamics.  
 CC -!- SUBUNIT: Homodimer or homotetramer.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=P30427-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P30427-2; Sequence=VSP\_005050;  
 CC Name=3;  
 CC IsoId=P30427-3; Sequence=VSP\_005051;  
 CC Name=4;  
 CC IsoId=P30427-4; Sequence=VSP\_005052;  
 CC -!- TISSUE SPECIFICITY: Widely expressed with highest expression in  
 skeletal muscle and lowest in thymus.  
 CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
 vimentin, desmin, GAP, cytokeratins, lamin B; whereas both the N-  
 and the C-terminus can bind integrin beta-4.  
 CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 33 plectin repeats.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.  
 CC -!- CAUTION: Isoform 4 is a fragment at the N-terminus.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).





FT DOMAIN 914 1239 COILED COIL (POTENTIAL).  
FT DOMAIN 1314 1418 COILED COIL (POTENTIAL).  
FT DOMAIN 1660 1765 DILUTE.  
FT NP BIND 163 170 ATP (POTENTIAL).  
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).  
FT MOD RES 1733 1733 PHOSPHORYLATION (POTENTIAL).  
SQ SEQUENCE 1828 AA; 211762 MW; 583DELC89AE36123 CRC64;

Query Match  
Best Local Similarity 3.1%; Score 295.5; DB 1; Length 1828;  
Matches 287; Conservative 231; Mismatches 538; Indels 427; Gaps 68;

QY 559 KLENEV--RNKELVAPTEBQSMVYLTKYRLTLKAAVELPFSQMRLPCSKVAVYN 616  
DB 410 KLFNWRVHVNQALHSAVKHSTIGVLDIYGFTEFINSPEQF-----CIN---YAN 458  
QY 617 KQALRIRSDNLHL-----DVVMQRTILELL-----CFNPLWLRLGLVVGKIQM 664  
DB 459 E---KLQQNMHVFKLEQEEYKMQLPWTLIDFYDNQPCINLIESKGLDLDLDECKM 515  
QY 665 QSNRDIVGLSTFNLRLFRNKC---BEQVYS-KAY-----TLT 698  
DB 516 PKGTDDTWAQKLYNTHL--NKCALFEKPRMSNAFIKHFPADKVEYQCEGFLEKNKDVF 573  
QY 699 EYEAETIKKHSLOKILPLPLDQAKOKRIVK-----HNPCLFVKKSP----- 741  
DB 574 EQIKVLKSKFK-----MLPELFQDDEKALSPSATSSTGRTPLTRVPVKTKGRPGQAK 629  
QY 742 -HKETKDILLRFSSELLANIGDITRELRRLGYVLQHRQTFDLDFDFAFNNAVLDLRGVR 800  
DB 630 EHKKTGVLQFNSLHL-----METLNATPHVYVCIKPNDFK-PFTFDEKRAVQ 679  
QY 801 LTRVVEVILLDDLTROLRVPA-----ISLQRI FNVKLALG----- 837  
DB 680 QLRACGYL-----ETIRISAGFPSPRWYQEFSSRYVLMKQKQVLDGRKOTQCNVLE 732  
QY 838 --ALGEANFOLGSDIAAQDIVDGHREKTLISLLWLIY--KFRSPKFAAAATVLOK---W 889  
DB 733 KLIDKOKYQFG-----KTKLFFRAGQVAVLEKLRADKLFAACIRIQKTIRGW 780  
QY 890 -WRRHWL-----HVVIQRRIR-----HKELMRHRRAATVIAVFRGHQMKYVVKLQKTE 937  
DB 781 LLRKRYLCMORAATVQRYRGYQARCYAKFLRRTKAATTKYRWYVYVRRYKI----- 836  
QY 938 RTQAAIILQKTRYLAKQILQYSY--HSLITIQ---RWRAQQLGRQHRQREVELREA 991  
DB 837 RRAATVILQSYLQYGLARNR-YRKILREKAVIIOKRVGW-----LARTHYKRTM---KA 888  
QY 992 AIFLQIRWRRLFAKLLAAAEATARLQSQKQAAASYIQMWRTYQLGRITQRIEFLR-- 1049  
DB 889 ILYLQCCFRNMARLKLKLEAR-----SVERYKKLHIG 924  
QY 1050 QRDILMFVQRRMRSKWMLQKQEFQOLKRAAINIQORWRAKLSMRKKNADYLAIRSVL 1109  
DB 925 MENKIMQLQKVD-----EQNDKYKCLMEKLTNLGVYNSETEKLNDVRLQLSBEEA 978  
QY 1110 KVQAYRATIQ-----WRIDENHYYSRLKNVICLOQLRALMKWREORENVLRLNASIL 1164  
DB 979 KVATGRVLSQEEIAKLKDLLEQTRSEKKS1---EERADKYQTEQLVSNLKEEN-TLL 1034  
QY 1165 VQKRYRQMOIQORNAYLRT--RKCIINQRRWRATIQMERERKNVILHLOTTTKRIQIK 1222  
DB 1035 KQKETLNHLNVAQAKENTETWERKL-----VETKQLELDLNDERLYQNLANEFSRLBER 1091  
QY 1223 F-RAKREMKQ-----RAEFLQI-----KKVT-- 1243  
DB 1092 YDDLKEEMTLNVPKPGHKRTDSTHSSNESSEYTFSEFAETDIAPTEBTEKKVPLD 1151  
QY 1244 ----LVVQKRRALLOMEKEQEVYHLHREVITKLORRFAOKSMFRFAKVGQAQAVSC 1299  
DB 1152 MSFLKQKRVTELGQEKQWQDLEDRKEEQVL-----RSKAGGERPQIRGABLGYES 1205  
QY 1300 LQMEWRNHL--LRKERNSTFLQROAATLQRR-----YRARIANMIKOLKSYA 1345

DB 1206 LK---RQELSENKKLKNELNLRKA---LSKSAPEVNAPGAPAYRVLMEQUTAVSEEL 1259  
QY 1346 QLKQAATITQTRYRAKAMQKVLYQKOREALI-----KVQRRYRGNL 1389  
DB 1260 DVRKEVLI---LRSQLVSQKEAIQPKDKNTWTDSTILLEDVQKMKDGETAQAVIGLK 1316  
QY 1390 EMRKQLEVYQKQOAVIRLQKWRISRD-----MELCKAGYRRIIRLSSLSIQKWR 1440  
DB 1317 ETNRLL-----SQLSQKRSHEAEALRGEIQSLKEENNQQQLLAQNLQPPPE 1367  
QY 1441 ATVQARRQBEIFLSTIRKVRILMQAFTRAVILMRQRRREFEMKERRAAVITQRRFRACML 1500  
DB 1368 ARTEASLQHEITRLTNTNLDLMEQ-----LEKQDKTVRKLLKQKLVFAKKIGELVEVGOM 1421  
QY 1501 KARQDVLIQSSV--LIVORKFRANSMKQARQEFVQLRTIAVHLOQKFRGKEL-MIEOR 1557  
DB 1422 ENISPGQIIDEPIRPVNIIPRKQDFQGMLEYKREDEQKLVKNLILLEKPRGVAVNLISGL 1481  
QY 1558 NCFQLLRCSMPGQARAGFM-ARKRFOALMTPMEMDLIRQKRAAKVIOF-----Y 1607  
DB 1482 PAVILFWC-----VRHADYLDQDKVRSLLTSTINSI-----KKVLKKGDDFTVVSF 1529  
QY 1608 WRGYLIR-----RRQKHQGLDIRKRIAQLRQEAQAVNSVRCKV-Q 1647  
DB 1530 WLSNTRCFHLHCLKQYSGEGFMKHTSRQNEHCLTNF--DLAEYRQ---VISDLAIQIYQ 1584  
QY 1648 EAVRELRG--RPIASDAL-----AVLSQDLRLSRTVPHILMMCMSEF 1686  
DB 1585 QLVKRVLENILQPMVMSGLEHEFIIQGVGVKPTGLKRTSSIADEGTYTILDSILRLQNSF 1644  
QY 1687 MSTFCYGINAQAIRSEVDKQILERCRIILN-----LARNSTTV----- 1726  
DB 1645 HSNVCQHGMDEPILIKQVKQMFYIVGAITLNNLLRDKMCSWKGMIQIRYNVSQLEEWLR 1704  
QY 1727 --NYFOEGG-----LVTIAQML--LRWCCKDSEIFNLCTLI 1759  
DB 1705 DKNLMNSGAKETLEPIQAAQLQVKKCTDDDAEACSMCNAL 1747

RESULT 8  
PLE1 HUMAN STANDARD; PRT; 4684 AA.  
AC Q15149; Q15148; Q16640;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Plectin 1 (PLTN) (FCN) (Hemidesmosomal protein 1) (HD1).  
GN PLEC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Placenta;  
RX MEDLINE=96210632; PubMed=8633055;  
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;  
RT "Human plectin: organization of the gene, sequence analysis, and  
RT chromosome localization (8q24).";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283 (1996).  
[2]  
RN SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
RX MEDLINE=96312447; PubMed=8698233;  
RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,  
RA Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,  
RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,  
RA Uitto J.;  
RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:  
RT cDNA cloning and genomic organization.";  
RL Genes Dev. 10:1724-1735 (1996).  
[3]  
RP VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.

MEDLINE=97049959; PubMed=8894687;  
 Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
 Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.,  
 "Homozygous deletion mutations in the plectin gene (PLEC1) in patients  
 with epidermolysis bullosa simplex associated with late-onset  
 muscular dystrophy.";  
 Hum. Mol. Genet. 5:1539-1546(1996).  
 [4]  
 VARIANT MD-EBS LEU-429 INS.  
 MEDLINE=21090821; PubMed=11159198;  
 Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker J.,  
 Mass W., Hametner R., Klaussegger A., Huber A., Pohla-Gubo G.,  
 Wiche G., Uitto J., Hintner H.,  
 "A compound heterozygous one amino-acid insertion/nonsense mutation in  
 the plectin gene causes epidermolysis bullosa simplex with plectin  
 deficiency.";  
 Am. J. Pathol. 158:617-625(2001).  
 [5]  
 VARIANT EBS1 TRP-2110.  
 MEDLINE=21841370; PubMed=11851880;  
 Koss-Harnes D., Hoeyheim B., Anton-Lamprecht I., Gjesti A.,  
 Joergensen R.S., Jahnson F.L., Olaisen B., Wiche G.,  
 Gedde-Dahl T. Jr.,  
 "A site-specific plectin mutation causes dominant epidermolysis  
 bullosa simplex Ogna: two identical de novo mutations.";  
 J. Invest. Dermatol. 118:87-93(2002).  
 CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and  
 CC microfilaments and anchors intermediate filaments to desmosomes or  
 CC hemidesmosomes. Could also bind muscle proteins such as actin to  
 CC membrane complexes in muscle. May be involved not only in the  
 CC crosslinking and stabilization of cytoskeletal intermediate  
 CC filaments network, but also in the regulation of their dynamics.  
 CC -!- SUBUNIT: Homodimer or homotetramer.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q15149-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q15149-2; Sequence=VSP\_005030;  
 CC Name=3;  
 CC IsoId=Q15149-3; Sequence=VSP\_005030, VSP\_005031;  
 CC -!- TISSUE SPECIFICITY: Widely expressed with highest levels in  
 CC muscle, heart, placenta and spinal cord.  
 CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
 CC vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N-  
 CC and the C-terminus can bind integrin beta-4.  
 CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa  
 CC simplex with muscular dystrophy (MD-EBS) [MIM:226670]; an  
 CC autosomal recessive disorder characterized by epidermal blister  
 CC formation at the level of the hemidesmosome and associated with  
 CC late-onset muscular dystrophy.  
 CC -!- DISEASE: Defects in PLEC1 are the cause of epidermolysis bullosa  
 CC simplex 1 (EBS1) [MIM:131950]; also called epidermolysis bullosa  
 CC simplex Ogna type. EBS1 is an autosomal dominant form of  
 CC epidermolysis bullosa simplex differentiated from the more  
 CC generalized form of Koeber [MIM:131900] and the localized form of  
 CC Weber and Cockayne [MIM:131800] by the occurrence of skin  
 CC bruising.  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 33 plectin repeats.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -!- SIMILARITY: Belongs to the plectin or cytolinker family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z54367; CA911196.1; -  
 DR EMBL; U53204; AAB05427.1; -  
 DR EMBL; U63610; AAB05428.1; -  
 DR EMBL; U63609; AAB05428.1; JOINED.  
 DR EMBL; X97053; CAAG5765.1; -  
 DR FIR; CS9404; A59404.  
 DR HSSP; Q01082; 1BKR.  
 DR Genew; HGNC:9069; PLEC1.  
 DR GK; Q15149; -  
 DR MIM; 601282; -  
 DR MIM; 236670; -  
 DR MIM; 131950; -  
 DR GO; GO:0008307; F:Structural constituent of muscle; TAS.  
 DR InterPro; IPR001589; Actbind\_actinin.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR001101; plectin repeat.  
 DR InterPro; IPR005326; S10 plectin\_N.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectin; 19.  
 DR Pfam; PF03501; S10 plectin; 1.  
 DR ProDom; PD006662; S10 plectin\_N; 1.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00250; PLEC; 34.  
 DR PROSITE; PS00019; ACTININ\_1; FALSE\_NEG.  
 DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
 KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
 KW Disease mutation.  
 FT DOMAIN 1 1470 GLOBULAR 1.  
 FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 2756 4684 GLOBULAR 2.  
 FT DOMAIN 175 400 ACTIN-BINDING.  
 FT DOMAIN 179 282 CH 1.  
 FT DOMAIN 295 397 CH 2.  
 FT REPEAT 645 710 SPECTRIN 1.  
 FT REPEAT 740 824 SPECTRIN 2.  
 FT REPEAT 837 930 SPECTRIN 3.  
 FT REPEAT 1315 1415 SPECTRIN 4.  
 FT DOMAIN 1469 2756 COILED COIL (POTENTIAL).  
 FT REPEAT 2826 2863 PLECTIN 1.  
 FT REPEAT 2864 2901 PLECTIN 2.  
 FT REPEAT 2902 2939 PLECTIN 3.  
 FT REPEAT 2940 2977 PLECTIN 4.  
 FT REPEAT 2981 3015 PLECTIN 5.  
 FT REPEAT 3116 3153 PLECTIN 6.  
 FT REPEAT 3154 3191 PLECTIN 7.  
 FT REPEAT 3192 3229 PLECTIN 8.  
 FT REPEAT 3230 3267 PLECTIN 9.  
 FT REPEAT 3268 3305 PLECTIN 10.  
 FT REPEAT 3306 3343 PLECTIN 11.  
 FT REPEAT 3485 3522 PLECTIN 12.  
 FT REPEAT 3523 3560 PLECTIN 13.  
 FT REPEAT 3598 3636 PLECTIN 14.  
 FT REPEAT 3599 3636 PLECTIN 15.  
 FT REPEAT 3640 3674 PLECTIN 16.  
 FT REPEAT 3820 3857 PLECTIN 17.  
 FT REPEAT 3858 3895 PLECTIN 18.  
 FT REPEAT 3896 3933 PLECTIN 19.  
 FT REPEAT 3934 3971 PLECTIN 20.  
 FT REPEAT 3975 4008 PLECTIN 21.  
 FT REPEAT 4061 4100 PLECTIN 22.  
 FT REPEAT 4101 4138 PLECTIN 23.  
 FT REPEAT 4139 4176 PLECTIN 24.  
 FT REPEAT 4177 4214 PLECTIN 25.  
 FT REPEAT 4218 4252 PLECTIN 26.  
 FT REPEAT 4265 4305 PLECTIN 27.  
 FT REPEAT 4319 4356 PLECTIN 28.  
 FT REPEAT 4408 4445 PLECTIN 29.  
 FT REPEAT 4446 4483 PLECTIN 30.  
 FT REPEAT 4484 4521 PLECTIN 31.



```

FT REPEAT 4522 4559 PLECTIN 32.
FT REPEAT 4560 4537 PLECTIN 33.
FT DOMAIN 4250 4300 BINDING TO INTERMEDIATE FILAMENTS
(BY SIMILARITY).
FT DOMAIN 4625 4640 4 X 4 AA TANDDEM REPEATS OF G-S-R-X.
FT MOD_RES 4539 4539 PHOSPHORYLATION (BY CDC2).
(BY SIMILARITY).
FT VARSPLIC 1 174 NYAGMLPRDQRLALVELFREGVAVAKDRPRSLHHPV
GVTVNQVAMASLRARGLVRETFAWCHFFWYLTNEGIAHL
ROYLHLPPEIVAASIQVRVRPVMYMPARTPHVQVQGPL
GSPPKRGLPTQORLYRKELEBYSPETPVVPTQRTLA
RQGPPEPAT -> MSGDEAVRAVEDVSNSSGSPSPGD
TLPWNLGKTQRRSGGAGNSVLDPAERAVIRIA (in
isoform 2 and isoform 3).
FT FT /FTid=VSP_005030.
FT VARSPLIC 409 412 Missing (in isoform 3).
FT /FTid=VSP_005031.
FT VARIANT 429 429 L -> LL (in MD-BBS).
FT /FTid=VAR_011336.
FT VARIANT 1003 1005 Missing (in MD-BBS).
FT /FTid=VAR_011337.
FT VARIANT 2110 2110 R -> W (in BBS1).
FT /FTid=VAR_015817.

Query Match 3.1%; Score 294; DB 1; Length 4684;
Best Local Similarity 19.7%; Pred. No. 6.5e-08;
Matches 326; Conservative 283; Mismatches 607; Indels 438; Gaps 71;

QY 249 VHATHTRALACIHE-----EEGSPRTP-TKSAIHLKRDIKLVGSPDKYSESMDLS 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 VYRLHERLVAITEYNLRKAGVAPATQVATLQSVORPELSDTIR-----YLDLL 666
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 LLSPTQKVAIQGS-----MPLNEM--KIRSTEQNRVYQEQIQIKADKLNSSSSSEASL 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 667 AWVEENQHRVDGAEGVDLPVSAQLGSHRGLQS--IEEQAKTERRASDEGQLSPATR 724
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 ACQOEFLNHSILLAQSFRNLHEVGRKSVKGVSPVNHKRSHEL-----SPSDAPSNE 411
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 GAYRDCI-----GRDLQYAKLLNSKARLSRLSLSHFSVAATKEL 766
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 412 LYRNETVAISPPKQORVEDTTLPRGAAPANASRSSAHAMP--HAQSKFKLAQTWSL 468
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 767 MWLINE-----KEEEVGFDWSDNTNMTAKESYSALMRELEKKEKIKELQN--- 814
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 469 MKKPTAPRKVR-DTSIQPSVKLYDSELYMOTCINPDPPAAFTTIDPFLASTMYLDEQAVD 527
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 815 ---AGDRLLREDHPARPVTESFOALQTC----- 840
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 528 RHQADFKKWLNALVSIPADLDLANNKIDVGLFNEVRNKE-----LVVAPTKEEQSMNY 582
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 841 -----W-SWMLQCCCEIAHLKENAYFOFSDVREAGQLQOEALRRKNSCDRS 891
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 LTKYRLTLRKAARELVFFSEQMRPLPCSKVAVVYKQALRIRSDRL-----HLDVVMQRTIL 639
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 892 ATVTRLEDLQDAQO-----EKQL--NEYKGLSLGLAKRAKAVVQLKPRHPAHPMGRPLP 945
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 640 ELLLCENPLNLRLGLVNVFGEKIOMQSNREDIVGLSTFFILNRLFRNKCEQRYSKAYTLTE 699
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 946 LLVAVCD---YKQVEVTVHKGDSCQ-----LVG-----PAQFSHWKVLSS 981
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 700 EYAEYTIKKHSLKILFLPFLDQAKQKRIK-----HNPCLFVKKSPHKEKTDIL---LR 751
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 982 SGSEA---AVPSVCLVPPNPQVQEAUTRLAQHQALVTLHQLHVDMSLLAQWSLR 1037
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 752 FSEELLANIGDTT-----RELRLGLVYLQ-HRQFTL-DEFD----- 785
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1038 RDVQIRSWSLATFTLKPQEQRALHSLHLYQAFRLRDSQDAGFGPEDILMAERYGS 1097
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 786 --YAFNNLAVDLRDG-----VRL-----TRVVEVILLRDDLRLQLRVP 821
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 CSHHVQQLLOSLEQAQESRCQRCISLKDRLQLQACETITVRLRLPLD-----KEP 1152
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 822 AISRLQIRFNKLA-----LGALEGEANFOLGGDIAAQDIVDGHREKTLISLLW 868
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

PLEI\_CRIGR

ID\_PLEI\_CRIGR

AC\_Q9U155;

DT 16-OCT-2001 (Rel. 40, Created)

STANDARD; PRT; 4473 AA.

1638 AVNS-----VRCKVQEAUVRFLGRF 1657

2038 ASDSELRQKGLVEDTLRQRRQVEEILAKASF 2071

1582 RFQALMTPENMDLIRQKRAKVIRYWRGVLIRRR-----QKHQGLDIDIRKIAQLRQEA 1637

1981 EATRLKTEABIALKEAEANERLRLAEDAEAFORRELEHQAQHKADIIEERLAQLR---K 2037

1527 KQARQEVQVLTAVHLQOKFKRKLMIQRNCFOLLRCMSPGFORARGFARK----- 1581

1924 ORLEASAGRPRELA---EEAARLRAEAEAKRQRLAEDDAARQARAEAEVLAELKLAATG 1980

1527 KQARQEVQVLTAVHLQOKFKRKLMIQRNCFOLLRCMSPGFORARGFARK----- 1581

1864 QORQLLEELARLQREAAATQKQLEAELAKVRAEMEVLLAKAKAEESRSTSEKSK 1923

1474 QORREPE-----MKRBAVVIRFRPARCAMLKARQDYQLIOSSVILVQKPF--ANRSM 1526

1804 EBAERARRRKGAEQAVRQRELAEBQELQORLASGTAGTQORLAEEQELIRLAETEOGE 1863

1435 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434

1744 REEAERAAQOQAAERAEAEERQERQWOLKANEALRLRLQAEVLQOKSLAQAEAEKOK 1803

1347 ---LQKWRATVQARRQREIFLSTIRKVRIM-----QAFIR---ATLLMR 1473

1684 EARNLCOAEVERARQVQVALETAQSAEAELOSKEASFATQALERSIQEHHVAQL 1743

1384 RYGNLEMRKQIIVYQKQOAVIRLOKWRNSIRD-MRLCKAGYRRIRLSLS----- 1434



16-OCT-2001 (Rel. 40, last sequence update)  
 15-MAR-2004 (Rel. 43, last annotation update)  
 Plectrin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated protein) (IFAP300) (fragment).  
 PLEC1.  
 Cricetus griseus (Chinese hamster).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.  
 NCBI TaxID=10029;  
 [1]  
 REPEAT SEQUENCE FROM N.A.  
 RX MEDLINE=20334248; PubMed=10873583;  
 RA Clubb B.H., Chou Y.-H., Herrmann H., Svitkina T.M., Borisy G.G., Goldman R.D.;  
 RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a hamster plectrin ortholog.";  
 RL Biochem. Biophys. Res. Commun. 273:183-187(2000).  
 RN [2]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=96215219; PubMed=8626512;  
 RA Malecz N., Folsner R., Stadler C., Wiche G.;  
 RT "Identification of plectin as a substrate of p34cdc2 kinase and mapping of a single phosphorylation site.";  
 RL J. Biol. Chem. 271:8203-8208(1996).  
 CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and microfilaments and anchors intermediate filaments to desmosomes or hemidesmosomes. May be involved not only in the crosslinking and stabilization of cytoskeletal intermediate filaments network, but also in the regulation of their dynamics.  
 CC -!- SUBUNIT: Homodimer or homotetramer.  
 CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N- and the C-terminus can bind integrin beta-4.  
 CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM INTERMEDIATE FILAMENTS DURING MITOSIS.  
 CC -!- SIMILARITY: Contains 1 actin-binding domain.  
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC -!- SIMILARITY: Contains 32 plectrin repeats.  
 CC -!- SIMILARITY: Contains 4 spectrin repeats.  
 CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 -----  
 DR EMBL; AF260753; AAF70372.1; -.  
 DR HSPG; Q01082; 1BKR.  
 DR InterPro; IPR001589; Actbind actinin.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR001101; Plectrin\_repeat.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00681; Plectrin; 20.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00250; PLEC; 34.  
 DR PROSITE; PS00019; ACTININ\_1; PARTIAL.  
 DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE; PS50021; CH; 2.  
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding; Phosphorylation.  
 KW NON\_TER  
 FT DOMAIN 1  
 FT DOMAIN 1  
 FT DOMAIN 1260 2544  
 FT DOMAIN 2545 4473  
 FT DOMAIN 4473 4499  
 FT DOMAIN 4499 4508  
 FT SPECTRIN 1.  
 FT REPEAT

FT REPEAT 529 613  
 REPEAT 626 719  
 FT SPECTRIN 3.  
 FT SPECTRIN 4.  
 FT DOMAIN 1204  
 FT DOMAIN 1258 2548  
 FT REPEAT 2615 2652  
 FT REPEAT 2653 2690  
 FT REPEAT 2691 2728  
 FT REPEAT 2729 2766  
 FT REPEAT 2770 2804  
 FT REPEAT 2905 2942  
 FT REPEAT 2943 2980  
 FT REPEAT 2981 3018  
 FT REPEAT 3019 3056  
 FT REPEAT 3057 3094  
 FT REPEAT 3274 3311  
 FT REPEAT 3312 3349  
 FT REPEAT 3350 3387  
 FT REPEAT 3388 3425  
 FT REPEAT 3429 3463  
 FT REPEAT 3609 3646  
 FT REPEAT 3647 3684  
 FT REPEAT 3685 3722  
 FT REPEAT 3723 3760  
 FT REPEAT 3761 3797  
 FT REPEAT 3800 3834  
 FT REPEAT 3852 3889  
 FT REPEAT 3890 3927  
 FT REPEAT 3928 3965  
 FT REPEAT 3966 4003  
 FT REPEAT 4007 4041  
 FT REPEAT 4043 4094  
 FT REPEAT 4197 4234  
 FT REPEAT 4235 4272  
 FT REPEAT 4273 4310  
 FT REPEAT 4311 4348  
 FT REPEAT 4349 4386  
 FT DOMAIN 4039 4089  
 FT DOMAIN 4414 4429  
 FT MOD\_RES 4328 4328  
 FT SEQUENCE 4473 AA; 509015 MW; E144615D3D61E3484 CRC64;  
 Query Match 3.0%; Score 287; DB 1; Length 4473;  
 Best Local Similarity 19.2%; Pred. No. 1.5e-07;  
 Matches 358; Conservative 317; Mismatches 736; Indels 456; Gaps 77;  
 QY 161 ETRGKQATIP--PDNLAAP-----TPTLKNVSKANDMRPRITDDLEDQATNKT 213  
 DB 960 EGLGKGVARLSAEAKVLAPEPSAPAPTIRSELE-----LTGLKLEQVRSLSAI 1009  
 QY 214 FDVKGSETINISLDTDCSRIDGQPHTPINKTTTIVHATHTR--ALACIHEEGSPSPR 270  
 DB 1010 Y-LEKLKTLISVIRSTQGAEEVLKAHERQLKEAQAVPATLQLEATKASLKKLRAQAQAQ 1068  
 QY 271 TPTKSAIHLKDKIKLVGSP-L-RKYSKMKSLSLSPOTKYAIQSGMNMNMKIRSIIE- 328  
 DB 1069 QPVFDTLRDELGAQEVGRLQQRHGERDVEVERWRKRVNQLERWQAVLAQIDVQREL 1128  
 QY 329 -----QNRVYQR-----QQIQIKAKDLNSSSSSEASLAGQQ---EFLVHS 366  
 DB 1129 EQLGRLRYRESADPLSLWLOAKRQROQIQAVPINSAQAREQJRKQKALLEIERHG 1189  
 QY 367 ETLAQSRSFNLHEVGKSVK-----PKQORVED-TTLPRSAAPANASARSSSAHA 451  
 DB 1189 EKVEECQFAKQYI--NAIKDYELQVTVKALEFVSPAKPKVQSGSESVIQEVDLR 1246  
 QY 411 SLRYNETVAISP-----PKQORVED-TTLPRSAAPANASARSSSAHA 451  
 DB 1247 TRYSELTTTTSQYIKFISLTLRMEERERLABQQRAERERLAEAGEAALAEKQRLAEHA 1306  
 QY 452 WPHAQSKFKLAQTMKMKPATPRKVRDTSIQPSVKLYDSEL--YMQTCINPDPAATTT 510  
 DB 1307 QAKAQA-ELEAQELQRMQEEVAREEAAVNAQQAQKRSIQBELQHLROSSAEIQAKAQ 1365

```
QY 511 IDPFLASTWYLD-----QAVDRHQADFKKWLNALVSIPLADLANN-KIDVGL 560
Db 1366 VEAABSRMKIEEIRVVRLQLETTTERQGGAEGLQALRAAEAEAKRQAQAEERL 1425
QY 561 FNEVRNK-----ELVV-----APTKEQSNVYITKRLT-----LRKAAVEL 598
Db 1426 RQVQVODESQRQAERLRYKAQAEAEKORALQALEELRLQAEAEERLROQAER 1495
QY 599 FFEQMRFLPC-----SKVAVVYNQALRIIRSDRLHLDVVMQRTILELLCFNPL 648
Db 1486 ARQVQVQALTAQSAEVELQSKRAEAKTAQLERTLOEHTVTVQLREKAE-----1537
QY 649 WRLGLVGVGKIOMQSNRDIVGLSTFILNRLFNKCEQRYSKAYITLTYEARTIKKH 708
Db 1538 --RRAQQQAERAREAREERL-----ERWLKANEALRLRLQAEVAQOR 1581
QY 709 SLOKILFLPLDQAKOKRIKVNPCLFVKKSPHKETKIDILLRFSSELLANIGDITRELR 768
Db 1582 SLAQ-----ADAEKQKE-----EAEREAR 1600
QY 769 RIGYVLQH--ROTFLDEFOYAFNNLAVLDLGDGVRITRVVVEVILLR-----DDLTRQLRVP 821
Db 1601 RRGKABEQAVRQELAEQLEKORQALAECTAQORLAAEQELIRLRAETEQGQORQLLEE 1660
QY 822 AISRLQR-----IFNVKILGALGEANFOLGGDIAODIVDGHREKTLISLLWQLI 871
Db 1661 ELARQREATAHKKQOELEAELA-KVRAEMEVILASKARAEESRSTSEKSKORLEABA 1719
QY 872 YKFRSPKPHAA-ATVILQKWRHRLHVLVITQIRIRKELMRHRAATVITQAVFRGHQMRKY 930
Db 1720 DRPFRLAEAEARLALAEAKR-----QRQLAEEADAARQAEAEVLT-----EKL 1765
QY 931 VKLFKTER--TQAAIILQKPTR-----RYLAQKQIYOSVHSIITITQWRRAQQLGRQR- 982
Db 1766 AAISEATRLKTAETALKEAENRERLRLAEDEAPQRR-----RUEQAAHLKA 1815
QY 983 --ORFVELREAA-----IFLQRIWRRLFAKLL-----AAAEFARIQ-----1018
Db 1816 DIEERLALQKASESELERQKGLVEDTLQRRQVEEILALKVSEKAAAGKAELELG 1875
QY 1019 --RSQKQAAASVYOMQWTYQLGRIOHFEFLRQDLIMFVORRWRKSWMLRQK-----1072
Db 1876 RIRSSAEDTWRSEKQAEQAAQRQALAEEOQRREAREERVQKSLAEEAEARQKAALE 1935
QY 1073 EFQOLKRAAINIQORWRKLSMEKCNADYALRLSSVLKVQAVRKATIQMRIDENHYISLR 1132
Db 1936 EVERLK-AKVEARLRLRABEQESARQLOLAQEAQKRIQAEKA-----1979
QY 1133 KNVICLQRLRAIMKMQRENYL-FLRN-----ASILVQKRYMRQMIQDRNAYLTRK 1187
Db 1980 -HAFVVOQREELQOTLQEQSMLERLGEAEAAARAAAEAEAREQ--AEREA-AQSRK 2035
QY 1188 CIINVR-----RWATLQMRERKNYLHQTITKRIQIKFRAKREMKQKQAEFLQ 1238
Db 2036 QVEAEERLKQSAEQAEQAAQAAAEKURKEABEQAEARRAQAEQAAALQKQADAEWEK 2095
QY 1239 LKKVTLVQRRRLQMRKQERYHLREVTIKLQRRFPAQKSM---RFWEAKRYGTQA 1295
Db 2096 HKKFA-----EQTLRQKAQVQELTTLA--LQLEETDH-QKSILDBELQRLKAEVTEA 2145
QY 1296 AVSCIQ-----MWRHHLRKRERNS-----FLQIRQAAITLQRRY 1331
Db 2146 AQRSQVBEELFSRVQMEELGKLRIRAEARNALILDKONTQRFLEAEFKMQVAABE 2205
QY 1332 RARLNMIKLQSAVLQKQAA--ITTQTRVRAKAMQKQVLYQ-----KOREAIIKVROR 1384
Db 2206 AARLSVAQ--EAARLQLAEDLQORALAEKLEKKAQVQAEATRLKAEAILQOQKE 2263
QY 1385 YRGNLEMRKQIEVYQORQAVIRLQKWRSI-----RDLRLCKAGYR-RIRLSLSIORK 1438
Db 2264 LAEQARRLQEDKEQMAQOQLVSETQGTLEVERQORQLEMSAEARLKLRAEWS-RAQ 2322
```

```
QY 1439 WRATVQARRRRE-----IFLSTIRKVRIMQAFIRATILMRQORREFEMKR-RAA 1486
Db 2323 ARAEEDAQRFRKQAEIEGKELHETELATQEKVTLVQ-----TLEIQROQSDHDAERLEA 2377
QY 1487 VTIQRRFR-----ARCAMLKARQDVLQIOSSVIL-----VQKFRANRSMKQARQEFVQ 1535
Db 2378 IAELEREKELKQBAKLLQKLS--EEMVTQOQELQTOALQKSLFKSDKSLLOREERFIE 2436
QY 1536 LRTIAVHLQOKFRKIRIMIEQRCNFQLLRCMFPQFQARAFGFWARKEFQALMTPEWMDLI 1595
Db 2437 -----QEKAKLEQLFQDEVAQAQQLR-----BEQQRQQRQMEQKQELV 2475
QY 1596 -----RQKRAKVIQRYWR--GYLIRRQKHQGL-----DIKRIALQPOEAKAV- 1639
Db 2476 ASMEEARROCEAEHAEVRRKQEBLQHLQROQOQEKILAEENQRLERLQRLLEEHRAL 2535
QY 1640 ---NSVRCKVQEAVERFL-RGRFTASDALAVLSQ-----LDRLSRVPHLLMWCSEFMS 1688
Db 2536 AHSEEIATAQAAAKALPNGR---DALDQPSMEVEPEHAFGLRQKVPAL-----2582
QY 1689 TFCYGINMAQIRSEVDKQILIERCSRILNLARVNSTVTTFQEGGLVTIAQMLLRWCDKD 1748
Db 2583 ----QLREAGILSABEQLRLEQGHITVAELSQRED--VRQYLQ-GRSSIAGLLKPTDEK 2635
QY 1749 SEIFNTL 1755
Db 2636 LSVYTAL 2642
```

RESULT 10  
HMW2 MYCGE  
ID HMW2 MYCGE STANDARD; PRT; 1805 AA.  
AC P47450;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory protein 2).  
DE HMW2 OR MG218.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=20997;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.H., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
EL Science 270:397-403(1995).  
RN (2)  
RP SEQUENCE OF 557-659 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random sequencing.";  
EL J. Bacteriol. 175:7918-7930(1993).  
CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY SIMILARITY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

QY	296	ESWKDLSLSPQTKY-----AIQCSMPNLNEMKIRSTEONYQC-----EQOIQIK 341
Dd	208	ENQKRLLSIYEVTYRELVSADNELQWYENIDNQIQFKHQVQTYTDELQSURKTKOLT 267
QY	342	AKDLNSSSESSEASLAGOOEFNFHSEILAOSSRFNLHEVGKSVKGSPVNPKRRSHL 401
Dd	268	KQELVDRESALRVKIDDADFINAR--LAB-----LDVAK-----QL 303
QY	402	SFSD--APSNESLYRNETVAISPCK-----KQVEDTTLPRSA-APANASRSSASHAWPH 454
Dd	304	SFOGITKQAQHVEDKLVALNKDEBLNTQKEAFFNROSALIDINLKQOENELFA-KH 362
QY	455	A--QSKKFCLAQTMSLMK----KPATPRKVDTSIQPSVKLYSSELYMQTCINDPPAAT 508
Dd	363	LEHQONEFEQOSDSLLKLTETXKALQHKNFFNESATK--SEELNQ---BRELFEKR 417
QY	509	TTIDPFL--ASTWYLD--QAVDRHOAFKKWLNALVISPADLDADLN-----NKIDV 557
Dd	418	REIDTLTQASLEYEHORESSQLDKQNBVKQHFONLEYAKKEDERNLLDDQKKVDS 477
QY	558	GKLFNEVRNKLVPAPTKEGOSMNYLUKYLETLRKAVALFFSEOMR-----LPC 608
Dd	478	EAlF--QLKEKVAQERKELEEL-YLVKKQKQDKENEL-LFFEKQLKHQADPFENELEA 532
QY	609	SKVAYYNKQALRTGS-----DRMLHDVVMQRTILELLICFNPLWLRLGLGVVFSGKIQ 663
Dd	533	KQBELFPAKGALE-RSTFIKLEDKCDLNTKAQQ-----IANEFSQ 571
QY	664	MQSNRDIVGLSTFTLINLFRN-KCEQRYSKAYTLTEYAETIKKHSLOKTILPLLPLDQ 722
Dd	572	LKTDKSADFELMWQNEYENLQOEKQKLQERYFYFERNAVLNRLLOQKEELL-----626
QY	723	AKQRIUVKHNPCLPVKKSPIHKETKDILLR-FSSIELAN-----IGDIETRELRRLGYVL 774
Dd	627	-----OQKETDQJTKFQEORLINOREHKELVASVEKQEIILKKL 668
QY	775	Q-HRQTFELDEFDYAFNNLA---VDLRDGVRLTRVVEVILLRDDTLRLRPALSRLORF 830
Dd	669	QDESQTSLN---ASKNLAREMAIKFKEKEIEATEKQLND-----VN 708
QY	831	NVKLALGALGEANPOLGD-----IAAQDIVDHREKTLISLLMOLIYKFRSPKPHAATV 885
Dd	709	NAEVIQADLAQNLQNSINQERSELQNAQRIADFENDSLKKL-----749
QY	886	LQKWWRZHLHWJLQRRIRHKELM-----RRHRAATVIQAVFRGHMRKYVKLPKTERTO 941
Dd	750	-----NEYLSLOKRLQELQETLEANQKQHSYQN-QAYPEG-----ELDKLNREKO 793
QY	942	AI--ILOKFTRYILAQKOLYQSYHSIIITIORWRAQQLGRHQORFVELREAAIFLORI 999
Dd	794	AFLNLRKQOTMEVDAILKORLSDKHOALNMOO-----AELDRKTH-----ELANNA-FLNHDA 843

RA Rodriguez O.C., Cheney R.E.;  
RT "Human myosin-Vc is a novel class V myosin expressed in epithelial  
RL cells.";  
CC J. Cell Sci. 115:991-1004(2002).  
CC -!- FUNCTION: May be involved in transferrin trafficking. Likely to  
CC power actin-based membrane trafficking in many physiologically  
CC crucial tissues.  
CC -!- TISSUE SPECIFICITY: Expressed chiefly in non-neuronal tissues.  
CC Particularly abundant in epithelial and glandular tissues  
CC including pancreas, prostate, mammary, stomach, colon and lung.  
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -!- SIMILARITY: Contains 1 dilute domain.  
CC -!- SIMILARITY: Contains 6 IQ domains.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF272390; AAF78783.1; .  
CC HSP; P08799; 1VOM.  
CC Genew; HGNC:7604; MYO5C.  
CC InterPro; IPR002710; DIL.  
CC InterPro; IPR000048; IQ\_region.  
CC InterPro; IPR001609; myosin\_head.  
CC Pfam; PF01843; DIL; 1.  
CC Pfam; PF00612; IQ; 5.  
CC PRINTS; PR00193; MYOSINHEAVY.  
CC ProDom; PD003376; DIL; 1.  
CC ProDom; PD000355; myosin\_head; 1.  
CC SMART; SM00015; IQ; 5.  
CC SMART; SM00242; MYSC; 1.  
CC PROSITE; PS00936; IQ; 4.  
CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Coiled coil; Polymorphism.  
FT DOMAIN 1 755 MYOSIN HEAD-LIKE.  
FT DOMAIN 756 779 IQ 1.  
FT DOMAIN 780 806 IQ 2.  
FT DOMAIN 807 829 IQ 3.  
FT DOMAIN 830 854 IQ 4.  
FT DOMAIN 855 884 IQ 5.  
FT DOMAIN 884 1351 COILED COIL (POTENTIAL).  
FT DOMAIN 1574 1679 DILUTE.  
FT VARIANT 522 522 P -> L.  
FT VARIANT 634 634 /FTID=VAR\_010646.  
FT VARIANT 634 634 L -> S.  
FT /FTID=VAR\_010647.  
SQ SEQUENCE 1742 AA; 202793 MW; 25DD3082A7EB5AAB CRC64;  
Query Match 3.0%; Score 281; DB 1; Length 1742;  
Best Local Similarity 17.7%; Pred. No. 8.9e-08;  
Matches 291; Conservative 274; Mismatches 515; Indels 560; Gaps 77;  
QY 415 NETVAISPPKQVDETTLPRAAPANASARSSAHAWPHAQSKFKLAQTMSLMKKEAT 474  
DB 382 SETV-----VKEMTPQAVNARDALAKKIYAHLPDFIVE---RINQALQFSGKQHT 429  
QY 475 PKVRDTSTQPSVKLYDSLYMTCTIN----- 501  
DB 430 FTGVLDIY---GFETFDVNSFQFCINYANEKLOQOQFNHVPKLEQSEYMKEDIPWTIID 486  
QY 502 -PDFFAATTIDPPLASTMYLDEQAVDRHQADPFKKWLNALVSIADLADLNKIDVGKL 560  
DB 487 FYDQNPVIDILIAKMGILDEECLEPLHGTD-ENWPQL-----YNNFVNRNPL 535  
QY 561 FNEVR--NKLTV---APTKERQSMNLYKYR-----LETIRKAAVEL-----PFSEQMR 605  
DB 536 FKPRMSNTSFVIQHFADKVEYKCEGFLEKNRDTVYDMLVELTRASKPHLCANFFQENPT 595

QY 606 LPSCKVAVYVYNKQALRIRSDRNHLHDVVMORTILELLCENPLWLRGLLEVVEGEKIOMQ 665  
DB 596 PP-----SP-----FGSMITVK 607  
QY 666 SNRDIVGLSTFILNRLFRNKCEQRYSKAYTL-----TEEYAEITKKHSLOKILFLLP 718  
DB 608 SAKQVIKPN---SKPRTTVGSKFRSSLYLLMETLNAITPHVYCRICKPNDEK---LP 658  
QY 719 FLDAQOKRIVKUNPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRRLGYVLQHRQ 778  
DB 659 F--EFDKSRIVQQLRACGVLET-----IRISAQYSP-----RW 690  
QY 779 TELDEPDYAFNNLAVDLRDGVRLTR-----VVEVILLRDLRQLRVPASIRLQR 828  
DB 691 TYIEFYS-----RYGILMTKQELSFSDKKEVKVVLHR-----723  
QY 829 IFNVKLALGALGAN-PQLGGDIAAQIDVDGHRKTLISLWQLIY--KFRSPKFAAATV 885  
DB 724 -----LIQDSNQYQFG-----NKKIFFRAGQVAYLEKLDLKLQSCVM 762  
QY 886 LOKWRRRHMLHVVIQRIIRHKELMRHRAATVIOAVPRGHQ-MRKYVKLFKTERTOAII 944  
DB 763 VQR-HMRGWL-----QKKFLRERRAALIIQQYFRGQQTVKKAITAVALKEAWAII 813  
QY 945 LQKFTREYLAQKOLYQSYH-SIITIQRWRA-----QQLGRQH-----R 982  
DB 814 IQHCRGYLV-RSLYQLIRMATITMQAYSRGFLARRRYKMLEHKAVALIKQVARAWLAR 872  
QY 983 QRFVELREAAFLQRIWRRLFAKL-----LAAETARLQSQQAAA 1026  
DB 873 RRFQSIIRFVLNITQYRVQRLQKLEQDNKENHGLVEKLTSLAALRAGDVEIKLEAF 932  
QY 1027 ASYIQMWRTYQ-----LGRIORH-EFLQRDILMFVQRRMRKSWMLEQR 1071  
DB 933 LEKAAATHRNVEYKGRYRDVAVEKLAQLQHNSELETQKEQI--QLKLEKTEELKEK 989  
QY 1072 KBFQOLKRAAINTQQRWRKLSMRKNADYALRSSLVKVQAY-----R 1115  
DB 990 MD-NLTKQLFDVQKEERQMLLEK-----SPELKTQDYEKQIQSKEBIKAKDE 1039  
QY 1116 KATIOMRIDRNHYYS--LRKNVICLOQLRAIMKREORENLYRLNASLTVOK----R 1168  
DB 1040 KMQLQHLVEGEHVTSDGLKAEVARLSQVKTISEFKEIE--LLOAQKIDVEKHVSQK 1096  
QY 1169 YRREQOM-----IQDRNAVLTREKCIINVQ-----RRWRATLQMRERKNYL 1210  
DB 1097 REMEEKMSSETTKOLLESYDIED---VRSLSVEDLEHLNEDGELMFA-----YE 1142  
QY 1211 HLQTTTTRIQIKFRKREMKKQRAEFLQAKVTLVQKRRALLQMRKEQVHLREVT 1270  
DB 1143 GLKKAATVLESHTFQSQDKCYEKEIEALNFKVHL-----SQEINHLOKL- 1186  
QY 1271 IKLQRRFPAOKSMRFMRKVRGTQAAVSCLOMHRNHLRKRERNFLQLRQAATLQRR 1330  
DB 1187 --FREENDINESIHVEVTRLTSENMMIPDFKQIIE--LEKQKDLIELRNEQA---EK 1238  
QY 1331 YRRLNMIKOLKSYAQLKQAAITITQRYRAKAKMQKVLYQKORAIL-KVQRRYRGNL 1389  
DB 1239 MKGKLELSNQLRHSREECT-----QRKALEAQNEIHTKEKEKLDIKIQEMQEAD 1290  
QY 1390 EMRKQIIEVYQKQ---QAVIRLQKWRISIR--DM--RLCKAGYRRIRLSLSI-----1435  
DB 1291 HLKQAFTESEVKNFQEAASRLTLENRDLREELMDKORVKKLQDVKTLSKTIGKAND 1350  
QY 1436 -----QRKWRATVQARRQREI-----FLSTI 1456  
DB 1351 VHSSSGPKGYLGMLOVKREDEAKLIQMLILDLPKRGVWNMMIPGLPAHILFMCVRYADSL 1410  
QY 1457 RKVELMQAFIRATL-----LMRQORREFFEMKRAAVIQRRFRFRCAMLKARQDYQIQS 1511  
DB 1411 NDANMLKSLMNSTINGIKQVVKHEHLEDFEM-----LSFWLNTCHUFINCLKQYS----1459  
QY 1512 SVILVQRFRFRANRSMKQAR-----QEFVQ-LRTIAVHLOQKFRGKRLMIFQRCNQL 1562



```

Db      948 KHK-----EELLAQQAELQELQEQATALASQDDTVQAKDLARQOHELELQON 999
QY      1135 VICLOQRRAIMKREORENYILRLNASILVQRYR--MRQOMTQDRNAVLTREKCIIN 1191
Db      1000 --AFNQASISLQKQRELTQNVKVLHGEL--KKREKLTLDRLAAREKQHKDAEIN 1055
QY      1192 VQRWRATQMRERKNYVHLQTTTR--IQIFRAKREMKQRA-----EFLQL 1239
Db      1056 -QR-----FKQFENEYADFQAKKRELQELNOTLRNLEQNASLKKRNOLTLDFALL 1107
QY      1240 KVVTLVQKRRALLQMRERQSYLHLREVTTIKLQRRPFAQKSMRPMRAKYRGTQ----- 1294
Db      1108 RKVQHTQNTNR---VQNTQIKFLEKKNQKQKASDEALQALILKLRSPASKLQIQR 1164
QY      1295 --AAVSCLOQWNRNHLKRENSFLQLQAAITLQRYVRAELNMKQLKSYAQ----- 1346
Db      1165 EALAIQKLEFDRDEQKQSEINNAKIQLEQFKLEKQNFDEAKQKQLEFQKQCRDLVEK 1224
QY      1347 --LKQAAITQTRYAKKAMQVVLVQKQREAILIKVQRRVGNLEMRKQIE----- 1396
Db      1225 RLKQKVLQKLSKSYLYTKRADLSQQQ-----LQHYANLLELKEKQGTAKRALDK 1278
QY      1397 ----VYQKQRAVIRLOKWRISIRDMRLCKAGYRIRLSSLSIOR-----KWRATVQARRQ 1448
Db      1279 KHRAYGKMAQFVSEL-----ROEKQLLSAQKQVDDKSRLLQKQNRH 1321
QY      1449 REIFLSTIRKVLQAFIRATLLMQORREPEMKRAAVIQRFRACMLKARQDYQL 1508
Db      1322 LQNLSETKKKQSQ-----LEHDINKFQRRKEAV----- 1351
QY      1509 IOSSVILVOKFRANRSMQARQEFVQLRTIAVHLQKQPRKRLMEQRCNPQLLRCSMP 1568
Db      1352 --SSIL-----NSHKKLKQKEGELQIL---QKLSLKKTLQEQE--FSKL----- 1389
QY      1569 GQPARARGMARKRFOALMTPPMMDLIQKRAKVIQRYWRGYLIRRRQKHQGLDI--- 1625
Db      1390 -YQREK-----LDQRTTLKHLRELKAQNBATAFKNRVLEIENY 1430
QY      1626 -RKRIQLRQEAQKAVNSVRCKQEAFLRGRFIASDA--LAVLSQDLRLSTVPHLLMW 1682
Db      1431 YKKEQLRLTEKSEFQNNKRLFEYFKIRNIEKKEAHKIVLEBQTKR-----HL--- 1483
QY      1683 CSEFMSTFCYGIMQAIRSEVQKLI-----ECSRIILNLARYNSTVNTVQOE 1731
Db      1484 -----VETEAVKLHLQKQSIQSKQELKEIKERVSRDISHNTKQREELNSLLHQ 1532
QY      1732 GGIIV--TIAQMLLRWCKDSEIPTNLTLLVWYFAHCPKRIIH--DYMTNPRAATMYRE 1787
Db      1533 NKLLQKNLAEREIRINNKQSLTLQKIQTAKQLSE--KEARLLKLEKMRVAVEQYQQA-E 1589
QY      1788 TKKLVARKEKMKQNAKRP--PQMTSGRYKSOKINFPTCPLSLPDLF 1832
Db      1590 ITRLKNRNLDEKNDKNHLPFLF-----KINGNDMNPYPYDPWF 1629

RESULT 13
TRHY_HUMAN
ID TRHY_HUMAN STANDARD; PRT; 1898 AA.
AC Q07283;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Trichohyalin.
GN TH OR TRHY OR THL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93280194; PubMed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,

```

```

RA      Steinert P.M.;
RT      "The structure of human trichohyalin. Potential multiple roles as a
RT      functional EF-hand-like calcium-binding protein, a cornified cell
RT      envelope precursor, and an intermediate filament-associated (cross-
RT      linking) protein."
RL      J. Biol. Chem. 268:12164-12176(1993).
RN      [2]
RP      SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX      MEDLINE=93315897; PubMed=7686953;
RA      O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;
RT      "Trichohyalin: a structural protein of hair, tongue, nail, and
RT      epidermis."
RL      J. Invest. Dermatol. 101:165S-71S(1993).
CC      1- FUNCTION: Intermediate filament-associated protein that associates
CC      in regular arrays with keratin intermediate filaments (KIF) of the
CC      inner root sheath cells of the hair follicle and the granular
CC      layer of the epidermis. It later becomes cross-linked to KIF by
CC      isodipeptide bonds. It may serve as scaffold protein, together
CC      with involucrin, in the organization of the cell envelope or even
CC      anchor the cell envelope to the KIF network. It may be involved in
CC      its own calcium-dependent postsynthetic processing during terminal
CC      differentiation.
CC      2- SUBUNIT: Monomer (Probable).
CC      3- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC      the inner root sheath (IRS) of hair follicles and medulla, and in
CC      the filiform papillae of dorsal tongue epithelium (Probable).
CC      4- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC      the epidermis.
CC      5- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC      calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC      entirely alpha-helical, configured as a series of peptide repeats
CC      of varying regularity, and are thought to form a single-stranded
CC      alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC      the most regular and may bind KIF directly by ionic interactions.
CC      Domains 5 and 7 are less well organized and may induce folds in
CC      the molecule. Domain 9 contains the C-terminus, conserved among
CC      different species.
CC      6- PTM: Substrate of transglutaminase. Some 200 arginines are
CC      probably converted to citrullines by peptidylarginine deiminase.
CC      7- SIMILARITY: In the N-terminal section; belongs to the S-100
CC      family.
CC      8- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; L09190; AAA65582.1; -.
DR      PIR; A45973; A45973.
DR      HSP; P02633; 4ICB.
DR      Genew; HGNC:11791; THH.
DR      MIM; 190370; -.
DR      GO; GO:0005856; C:cytoskeleton; NAS.
DR      GO; GO:0005509; F:calcium ion binding; TAS.
DR      InterPro; IPR001751; CaBP_S100.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR002017; Spectrin.
DR      Pfam; PF000036; ehand; 1.
DR      Pfam; PF01023; S_100; 1.
DR      PRODOM; PD003407; CaBP_S100; 1.
DR      PROSITE; PS000018; EF_HAND; 1.
DR      PROSITE; PS00303; S100_CaBP; 1.
DR      Keratinization; Calcium-binding; Repeat; Citrullination.
FT      DOMAIN 1 91 S-100 LIKE.
FT      CA_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).
FT      CA_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).
FT      DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF
FT      R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT      REPEAT 314 326 1-1 (APPROXIMATE).

```











```
FT DOMAIN 789 813 IQ 2.
FT DOMAIN 814 836 IQ 3.
FT DOMAIN 837 861 IQ 4.
FT DOMAIN 862 884 IQ 5.
FT DOMAIN 885 913 IQ 6.
FT DOMAIN 914 1237 COILED COIL (POTENTIAL).
FT DOMAIN 1314 1443 COILED COIL (POTENTIAL).
FT DOMAIN 1685 1790 DILUTE.
FT NP BIND 163 170 ATP (POTENTIAL).
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
FT MOD_RES 1758 1758 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 1853 AA; 215594 MW; 503E93D48CA6B766 CRC64;

Query Match
Best Local Similarity 18.7%; Score 271.5; DB 1; Length 1853;
Matches 290; Conservative 250; Mismatches 564; Indels 449; Gaps 70;

QY 518 TWYLDQAVDRHQ-----ADPKWLNALVSIIPADLDADLNKNDVGLKFNVEARN 566
DB 358 TIFCDLMGVDYBEMCHWLCHRKLATATETVYKPSIKLOATNARDALAKHIYAKLFNWIVD 417
QY 567 --KELVAPTEKESQSNYLTKYELETLRKAAVELFSEQMRLPCSKVAVVNVKQALRIS 624
DB 418 HYNQALHSAVKQHSFGLVDIYGFETFEINSFEQF-----CIN--YANB---KJQQ 463
QY 625 DRNLHL-----DVMQRTILELL-----CNPLWRLGLLEVVFGEKIQMSNRDVG 672
DB 464 QFNHMFVKLEQEBYMEQIWPWTLIDFYDQPCINLIESKIGILDLDLDECKMPKGTDDTW 523
QY 673 LSTFILNRLFRNCK-----EQRYS-KAY-----TLTHEYAEATIK 706
DB 524 AQKLYNTHL--NKCALFERPRMSNKAIIKHFAKVQEYQCEGFLEKNKQDVFEEQIKVLK 581
QY 707 KHSLOKILFLPLDQAKOKRIVK-----HNPCLFVVKSP-----HKETKDI 748
DB 582 SKFK-----MLPELFODDEKAISPTSATSGRTPLTRVPVKTPGPGQTAKHEKHTVGH 637
QY 749 LRFSSSELLANIGDITRELRRLGYVLQHRQTFLDDEPDYAFNNLAVDLRDGVRLTR---VV 805
DB 638 QFNSLHL-----METLNATPHVYVCIKPNDFKF-PPTFDEKRAVQQLRACGVL 687
QY 806 EVILL--RDLTRQLRVPALSRQIRFNVKLAG-----ALGEANFOLGGDI 850
DB 688 ETIRISARGFPFRWTYQEPFSRYRVLMKQKVDLGRKQTKCNVLEKLIIDKQKYQFG---744
QY 851 AAQDIVDGHREKTLSSLWOLIV--KFRSPKFAAAATVLOK---W-WRRHWL-----HVV 898
DB 745 -----KTKIFFAGQVAVLEKLRADKLRAACIRIQKTIRGWLRLKRYLCMQRAAIT 795
QY 899 IQRRIP-----HKELMRHRAATVIQAVPRGHOMRKYVKLFKFTERTOAAILQKFTRRY 952
DB 796 VQYVRGYQARCYAKFLRRTKAATTIQYRMVYVRRYKI---RRAATIVIQSYLRGY 851
QY 953 LAQ--KQIYQSYHSIIITIQ---RWRAQOLGROHRORFVELREAAIFLQRIWRRRLPAK 1006
DB 852 LTENRYKILREYKAVI-IQKRVGW---LARTHVKRTM---KATVYLQCCFR-----898
QY 1007 KLLAAAEATLARLQRSQKQAAASYIQWRTYQLGRIQREHFLR--QRDLIMFVQRRMRGK 1064
DB 899 -MMAKEDVKKLEARS-----VERYKKLHIGNENKIMQLQKVD--937
QY 1065 WSNLQORKEFOQLKRAAINIQWRRAKLSNRKCNADYLALRSSVLKQAYRKATIO----1120
DB 938 ----EQNKDYKCLMEKLTNLEGVYNSETEKLRNDVERLQLSBEAKVATCRVLSLQEEIA 993
QY 1121 -MEIDRNHYYSLRKNVICIQRLRAIMKWRBQRENVYLRNANLILVOKYRMQMIQDR 1179
DB 994 KLRKOLEQTRSEKKS1---EERADKYQETDQDVSNLKEEN-TLLQKEKETLNHRIVEQA 1049
QY 1180 NAYLRT--RKCIINVQRRWRATLQMRERKNYIHLQTTTKRIQIKF-RAKREMKQ----1232
DB 1050 KEMTETMERKL---VEETKQLELDLNDERLRYQNLNANFESRLEERYDDLKEEMTLMNVP 1106
```

```
QY 1233 -----RAEFLQL-----KKVT-----LVQKRRRALL 1254
DB 1107 KPCHKRTDSTHSSNESEYTFSEFAETDIAPRTEPIEKVKPLDMSLFLKQKRVTELE 1166
QY 1255 QMRKEROYVILHREVITIKLQRRPFAQKSMRFRKRYGTCQAASVCLQMRWRNL--LRKR 1312
DB 1167 QEKQLMODBLDRKE-----EQVFRSKAKBEERPOIRGAELYESLK--RQSESENKK 1217
QY 1313 ERNSFLQROAAITLQRR-----YIARLNMIKQLKSYAOLKQAAITIQRYRA 1360
DB 1218 LKNELNELRKA--LSEKSAPEWAPAGAPAYRVLMQELTSVSEELDVRKEEVL---LRS 1271
QY 1361 KKAMQKQVVLVYQKQREAI-----KVORRYRGNLBMRKQIEVYQKQQA 1404
DB 1272 QLVSQKEAIQKDDKNTMTDSTILLIEDVQKMDKGBIAQAYIGLKETNRLLE-----1323
QY 1405 VIRLQKWNISIRD-----MRLCKAGYRIRLSSLSIQKWRATVQARRQREIFLST 1455
DB 1324 -SQLOSKKSHENEAELRGEIQSLKEENRQOQLAQNLOLPPEARIASLQHEITRLT 1382
QY 1456 IRKVLMOAF-----IRATL-----LMRQORREFEMKRAAVVIQRRFRAR 1496
DB 1383 NENLIFEELYADDPKYQSYRISLYKGMIDLMEOLEKQDQKTVKLLKQKLVFAKKIGELE 1442
QY 1497 CAMLKARQDYQLIQSSV--ILVQRKFRANRSMKQARQEFVQLRTTIAVHLQOKFRGKRLMI 1554
DB 1443 VQGMENISPGQIIDEPIRPVNI PRKEKDFQGMLEYKREDEQKLVKNLILELPRGVAVNL 1502
QY 1555 EQNRCFQLLRCSMPQOARARGFMA-----KRFQALMTPEMMDLIRQKAAKVIQR 1606
DB 1503 -----IPGLFAYILFWCVRHADYLNDDQKVRSLTSTINSI-----KKVLKK 1544
QY 1607 -----YMRGYLIR-----RRQKHGOLLDIRKIRIAOLRQEAKA 1638
DB 1545 RGDDETVTSFMLSNTCRFLHCLKQYSGEGFMKHTSRQNEHCLTNF--DLAERYQ---V 1599
QY 1639 VNSVRCKV-OEAVRFLRG--RFIASDAL-----AVLSQDLRLSRTV 1676
DB 1600 LSDLAIQIYQQLVRVLENIQPMIVSGMLEHETIQGVSGVKPTGLRKRTSSIADEGTYL 1659
QY 1677 PHLLMWCSEFMSTFCYGINAQAIARSEVDKOLIERCSRIILN-----LARY 1721
DB 1660 DSILRQLNSFHSVWMOHGMDBELIKQVVKOMFYIVGAIITNNLLLRKDKMCSWSKGMQIRY 1719
QY 1722 NSTTV-----NTFOEGG-----LVITIAQML--LRWCDKDSIEFNTLCTLI 1759
DB 1720 NVSQLEEWLRDKNLMNSGAKETLEPLIQAAQLLVKKKTTDDDAEAICSMCNAL 1772
```

Search completed: September 13, 2004, 17:41:15  
Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:37:14 ; Search time 103 Seconds  
(without alignments)  
5700.771 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514  
Sequence: 1 MELVNSPVLEVACKETLQLI.....FISSVAFDTTLCKLQIDMF 1861

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.25.\*  
1: sp arChaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9514	100.0	1861	5	O01401	O01401 drosophila
2	9482	99.7	1861	5	Q9VC45	Q9vc45 drosophila
3	4200	44.1	877	5	Q8X666	Q8x666 drosophila
4	1374	14.4	3122	11	Q8CJ27	Q8cj27 mus musculus
5	1360	14.3	3477	4	Q8IZT6	Q8izt6 homo sapien
6	868	9.1	1142	4	Q8N4D1	Q8n4d1 homo sapien
7	733.5	7.7	825	4	Q8IUL2	Q8iul2 homo sapien
8	733.5	7.7	826	4	Q8IZJ8	Q8izj8 homo sapien
9	688	7.2	882	4	Q8IZJ9	Q8izj9 homo sapien
10	408.5	4.3	726	4	Q9NWS1	Q9nws1 homo sapien
11	370.5	3.9	1110	10	Q84Z28	Q84z28 oryza sativ
12	361.5	3.8	1186	5	O17666	O17666 caenorhabdi
13	324.5	3.4	1088	10	O49705	O49705 arabidopsis
14	314	3.3	1909	5	Q25893	Q25893 plasmodium
15	300	3.2	1374	10	Q9SMY9	Q9sm9 arabidopsis
16	290.5	3.1	285	4	Q9NVT6	Q9nvt6 homo sapien

17	281	3.0	575	11	O88482	O88482 mus musculu
18	272.5	2.9	2473	11	Q9QZ84	Q9qz84 mus musculu
19	271	2.8	1705	5	Q9U0S7	Q9u0s7 mytilus gal
20	270	2.8	2009	11	O8CJ40	O8cj40 mus musculu
21	268.5	2.8	2245	5	Q86A36	Q86a36 dictyosteli
22	267.5	2.8	1596	5	Q8IJ44	Q8ij44 plasmodium
23	262.5	2.8	1849	5	Q9U487	Q9u487 loligo peal
24	261.5	2.7	446	5	Q9GYC6	Q9gyc6 leishmania
25	259.5	2.7	1919	4	Q7Z2L4	Q7z2l4 homo sapien
26	257.5	2.7	1800	5	Q8MKM1	Q8mkm1 drosophila
27	254.5	2.7	1837	3	Q74424	Q74424 schizosacch
28	254	2.7	3267	5	Q8IBZ0	Q8ibz0 plasmodium
29	252.5	2.7	1845	11	Q7TQL2	Q7tql2 mus musculu
30	252.5	2.7	2115	4	Q14980	Q14980 homo sapien
31	252	2.6	1764	5	Q9N9U8	Q9n9u8 leishmania
32	251.5	2.6	1792	5	Q97417	Q97417 drosophila
33	251.5	2.6	1792	5	Q9V4M6	Q9v4m6 drosophila
34	251.5	2.6	2094	11	O8OY35	O8oy35 mus musculu
35	250.5	2.6	1416	4	Q9BZF9	Q9bzf9 homo sapien
36	250.5	2.6	1416	4	Q9HCL1	Q9hcl1 homo sapien
37	250.5	2.6	1992	11	Q8OZE6	Q8oze6 mus musculu
38	250.5	2.6	1995	4	Q7Z406	Q7z406 homo sapien
39	250	2.6	2048	5	Q9VCD1	Q9vcd1 drosophila
40	250	2.6	2101	4	Q14981	Q14981 homo sapien
41	248.5	2.6	1302	4	Q7Z2L3	Q7z2l3 homo sapien
42	245.5	2.6	1449	5	Q9BG87	Q9bg87 bos taurus
43	245	2.6	1401	5	Q7YU40	Q7yu40 drosophila
44	242	2.5	3616	13	Q9W6V0	Q9w6v0 gallus gall
45	239.5	2.5	1489	5	Q8T8Q5	Q8t8q5 drosophila

## ALIGNMENTS

RESULT 1  
O01401 ID O01401 PRELIMINARY; PRT; 1861 AA.  
AC O01401;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Microtubule associated protein.  
GN ASP OR CG6875.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON R;  
RA Saunders R.D.C., Avides M.C., Howard T.I.A., Gonzalez C.,  
RA Glover D.M.G.; 0.0-0.0).  
RL J. Cell Biol. 0.0-0.0).  
DR EMBL; U95171; AAB51540.1; --.  
DR PIR; T13845; T13845.  
DR FlyBase; FBgn0000140; asp.  
DR GO; GO:0005875; C:microtubule associated complex; IDA.  
DR GO; GO:0008017; F:microtubule binding; IDA.  
DR InterPro; IPR001715; Galponin-like.  
DR InterPro; IPR000048; IQ\_region.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00612; IQ; 21.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00015; IQ; 5.  
DR PROSITE; PS00021; CH; 1.  
DR PROSITE; PS00096; IQ; 5.  
SQ SEQUENCE 1861 AA; 219558 MW; 0A943D5B335EFF45 CRC64;

Query Match 100.0%; Score 9514; DB 5; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELVWSPLYEVA CKETLQOLIDNRNFRKEVMI ILKSKSNQPVKNPRKPTVGKTLQLKSP 60  
Db 1 MELVWSPLYEVA CKETLQOLIDNRNFRKEVMI ILKSKSNQPVKNPRKPTVGKTLQLKSP 60  
QY 61 GAGTKMSVVSAA VQOKRMSAAAPSPKOTWRVTAPSRPAANAHPPQAPLVEKNVYKT 120  
Db 61 GAGTKMSVVSAA VQOKRMSAAAPSPKOTWRVTAPSRPAANAHPPQAPLVEKNVYKT 120  
QY 121 POEPVVISQPRSLKENLSFMTPGNLLDVINDNLRFPLTETRGKQATIFPDNLAAPT 180  
Db 121 POEPVVISQPRSLKENLSFMTPGNLLDVINDNLRFPLTETRGKQATIFPDNLAAPT 180  
QY 181 PTLKGNVKS CANDMRPRITPDDEDOPATNKFDVKHSETIMISLDTDCSIDGPH 240  
Db 181 PTLKGNVKS CANDMRPRITPDDEDOPATNKFDVKHSETIMISLDTDCSIDGPH 240  
QY 241 PLNKTTTTIVHATHTRALACIHEEBGSPPTPTKSAIHDLDKRDIKLVGSPLRKYSMD 300  
Db 241 PLNKTTTTIVHATHTRALACIHEEBGSPPTPTKSAIHDLDKRDIKLVGSPLRKYSMD 300  
QY 301 LSLSPOTKYAI QOSMPNLENMKIRSTEONRYQBOOI QIKAKDLNSSSSSEASLAGOE 360  
Db 301 LSLSPOTKYAI QOSMPNLENMKIRSTEONRYQBOOI QIKAKDLNSSSSSEASLAGOE 360  
QY 361 FLFNHSEILAQSSRFNLHEVGRKSVKSPVKNPHKRSHLSFSDAPSNESLYNETVAI 420  
Db 361 FLFNHSEILAQSSRFNLHEVGRKSVKSPVKNPHKRSHLSFSDAPSNESLYNETVAI 420  
QY 421 SPKKQVVEDTTLPRSAAPANASARSSSAHAWPHAKSKPKLAQTMSLMKPKPATPRKVRD 480  
Db 421 SPKKQVVEDTTLPRSAAPANASARSSSAHAWPHAKSKPKLAQTMSLMKPKPATPRKVRD 480  
QY 481 TSIOPSVKLYDSSELYMOTCINPDPEAATTTIDPELASTMYLDEQAVDRHOADPKKMINAL 540  
Db 481 TSIOPSVKLYDSSELYMOTCINPDPEAATTTIDPELASTMYLDEQAVDRHOADPKKMINAL 540  
QY 541 VSPADLDADLNKIDYVKLFNEVRNKLVAAPTKEEQSMNYLT KYRLETLRKA AVELFF 600  
Db 541 VSPADLDADLNKIDYVKLFNEVRNKLVAAPTKEEQSMNYLT KYRLETLRKA AVELFF 600  
QY 601 SEQMLPCSKVAVYVKNQALRISDRNHLVDVWQRTILELLICFNPMLRLGLEVVFG 660  
Db 601 SEQMLPCSKVAVYVKNQALRISDRNHLVDVWQRTILELLICFNPMLRLGLEVVFG 660  
QY 661 KI QMSNRD IVGLSTFIINLFRNKCEQRYSAYTLTEEYAEYTIKKHSLOKILFLLPFL 720  
Db 661 KI QMSNRD IVGLSTFIINLFRNKCEQRYSAYTLTEEYAEYTIKKHSLOKILFLLPFL 720  
QY 721 DOAKOKRIVKHNPCLFVKKSPHKETKDI LLRPSSELLANIGDITREIRLRYGVYLOHROT 780  
Db 721 DOAKOKRIVKHNPCLFVKKSPHKETKDI LLRPSSELLANIGDITREIRLRYGVYLOHROT 780  
QY 781 LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTROLRVPALSRQIFNVKALCALG 840  
Db 781 LDEFDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTROLRVPALSRQIFNVKALCALG 840  
QY 841 EAMFOLGDDIAODIVDGHREKTLISLWOLYIYFRSPKFAAATVI QKWRHRLHVVQ 900  
Db 841 EAMFOLGDDIAODIVDGHREKTLISLWOLYIYFRSPKFAAATVI QKWRHRLHVVQ 900  
QY 901 RRIHREKELMRHRAATVIOAVFQGHQMKVKVFLKFKTERTOAAIIOKFTTXYLAOKLYQ 960  
Db 901 RRIHREKELMRHRAATVIOAVFQGHQMKVKVFLKFKTERTOAAIIOKFTTXYLAOKLYQ 960  
QY 961 SYHSIIITQWRRAQQLGROHROFVELREAAIFLQRIWRRRLFAKLLAAAFARLORS 1020  
Db 961 SYHSIIITQWRRAQQLGROHROFVELREAAIFLQRIWRRRLFAKLLAAAFARLORS 1020  
QY 1021 QKQOAAASYIQOMWRTYQLGRIORHBEFLORDLIMFVQRMRSKWSMLERKEFOOLKKA 1080  
Db 1021 QKQOAAASYIQOMWRTYQLGRIORHBEFLORDLIMFVQRMRSKWSMLERKEFOOLKKA 1080  
QY 1081 AINTQORWRAKLSMRKCNADYLALRSVLKQVAYRKATI QMRIDRNHYYSLRKNVCILOQ 1140

Db 1081 AINTQORWRAKLSMRKCNADYLALRSVLKQVAYRKATI QMRIDRNHYYSLRKNVCILOQ 1140  
QY 1141 RLRAIWKREORENVLRLNASILVQKYRMRQOMI QDRNAYLTRKCI INVQRWRATL 1200  
Db 1141 RLRAIWKREORENVLRLNASILVQKYRMRQOMI QDRNAYLTRKCI INVQRWRATL 1200  
QY 1201 QMRERKNYLHIQTTTKRIQIKFRAKREMKQORAEFFLOLKKVTLVVOKRRBALLOMRK 1260  
Db 1201 QMRERKNYLHIQTTTKRIQIKFRAKREMKQORAEFFLOLKKVTLVVOKRRBALLOMRK 1260  
QY 1261 QBYLHLREVTTIKLORRPHAQSMRFRMRAKYRGTOAAVSCLOWHNRHLLRKRNSFLQ 1320  
Db 1261 QBYLHLREVTTIKLORRPHAQSMRFRMRAKYRGTOAAVSCLOWHNRHLLRKRNSFLQ 1320  
QY 1321 RQAATLQORRYARLANMIKOLKSYAQLKQAAITITQTRYAKKAMQOVVLYQKOREAIK 1380  
Db 1321 RQAATLQORRYARLANMIKOLKSYAQLKQAAITITQTRYAKKAMQOVVLYQKOREAIK 1380  
QY 1381 VQRRYRGNLEWRKOIEVYOKORQOAVIRLOKWRSTIRDMELCKAGYRIRLSLSIORK 1440  
Db 1381 VQRRYRGNLEWRKOIEVYOKORQOAVIRLOKWRSTIRDMELCKAGYRIRLSLSIORK 1440  
QY 1441 ATVOARROREIFLSTIRKVRMLQAFIRATILMRQORREFEMKRAA VVIQRRFRACML 1500  
Db 1441 ATVOARROREIFLSTIRKVRMLQAFIRATILMRQORREFEMKRAA VVIQRRFRACML 1500  
QY 1501 KARQDYOLI QSSVILVQRYFRANRSMKQARQEFVOLRTIAVHLOOKFRGKLMIEORNC 1560  
Db 1501 KARQDYOLI QSSVILVQRYFRANRSMKQARQEFVOLRTIAVHLOOKFRGKLMIEORNC 1560  
QY 1561 QLLRCSMPFGQARAGFMARKFOALMTPMMDLIROKRAAKVIQRYWRGYLIRRRQKHQ 1620  
Db 1561 QLLRCSMPFGQARAGFMARKFOALMTPMMDLIROKRAAKVIQRYWRGYLIRRRQKHQ 1620  
QY 1621 GLLDTRKRIAQLOREKAVNSVRCKVQBAVFLRGRFTIASDALAVLSOLDRLSRTVPHLL 1680  
Db 1621 GLLDTRKRIAQLOREKAVNSVRCKVQBAVFLRGRFTIASDALAVLSOLDRLSRTVPHLL 1680  
QY 1681 MWCSEFMSTFCYGINMAQAIRSEVDKOLIERCSRILNLARVNSTVTNTFOGGVTTIAQM 1740  
Db 1681 MWCSEFMSTFCYGINMAQAIRSEVDKOLIERCSRILNLARVNSTVTNTFOGGVTTIAQM 1740  
QY 1741 LLRWCDKDSIEFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYVRETCKLVARKEKMKQ 1800  
Db 1741 LLRWCDKDSIEFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYVRETCKLVARKEKMKQ 1800  
QY 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPDGFIIRSPYTFISSVYAFDTILCKLOIDM 1860  
Db 1801 NARKPPMTSGRYKSOKINFTPCSLPSLEPDGFIIRSPYTFISSVYAFDTILCKLOIDM 1860  
QY 1861 F 1861  
Db 1861 F 1861  
RESULT 2  
QYVC45 PRELIMINARY; PRT; 1861 AA.  
AC QYVC45;  
DT 01-MAY-2000 (TremBLrel. 13, Created)  
DT 01-MAY-2000 (TremBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE CG6875-PA.  
GN ASP OR CG6875.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2019606; PubMed=10731132;



```

Db 781 LDFDYAFNNLAVLDLGRVLRTRVMEVILLRDDLTRQLRVPAISRILQRIENVKLALGALG 840
Qy 841 EANFOLGGDIAACDIDVGHREKTLISLLWOLIVKPSPEKHAATAVLQKWRHSHLHVLIQ 900
Db 841 EANFOLGGDIAACDIDVGHREKTLISLLWOLIVKPSPEKHAATAVLQKWRHSHLHVLIQ 900
Qy 901 RRIHKLMMRRHRAATVIAVFRGHQMRKYVKLFKTERTOAAIIILQKFTRRYLAQKQLYQ 960
Db 901 RRIHKLMMRRHRAATVIAVFRGHQMRKYVKLFKTERTOAAIIILQKFTRRYLAQKQLYQ 960
Qy 961 SYHSIIITQWRRAOQLGHQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAEARLORS 1020
Db 961 SYHSIIITQWRRAOQLGHQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAEARLORS 1020
Qy 1021 QKQAAAASYIQOWRTYQGRIGRORHEFLRORDLIMFVQRMESKWSMLBQRFQOLKKA 1080
Db 1021 QKQAAAASYIQOWRTYQGRIGRIGRORHEFLRORDLIMFVQRMESKWSMLBQRFQOLKKA 1080
Qy 1081 AINIQRWRRAKLSMRKCNADYLALRSSVLKQAYRKATIQMRIDRNHYYSLEKNVICLOQ 1140
Db 1081 AINIQRWRRAKLSMRKCNADYLALRSSVLKQAYRKATIQMRIDRNHYYSLEKNVICLOQ 1140
Qy 1141 ELRAITMKREQENYILRNASTIIVQKRYMRQOMIDBNAYLTKCTIINVQRRWATL 1200
Db 1141 ELRAITMKREQENYILRNASTIIVQKRYMRQOMIDBNAYLTKCTIINVQRRWATL 1200
Qy 1201 QMRREKNYLHLQTTTKRIQIKFRAKREMKQABFLQIKKVTLVQKRRRALLQMKER 1260
Db 1201 QMRREKNYLHLQTTTKRIQIKFRAKREMKQABFLQIKKVTLVQKRRRALLQMKER 1260
Qy 1261 QBYLHLREVITIKQRFFHAQSMRPMRAKYRGTOAAVSCLOQWHNHLKREKNSFLQL 1320
Db 1261 QBYLHLREVITIKQRFFHAQSMRPMRAKYRGTOAAVSCLOQWHNHLKREKNSFLQL 1320
Qy 1321 FOAATLQRRYRANLMIKLSYALQKALITQITRYRAKAMQKQVVLQKQREALIK 1380
Db 1321 FOAATLQRRYRANLMIKLSYALQKALITQITRYRAKAMQKQVVLQKQREALIK 1380
Qy 1381 VQRRYRGNLEMRKQLEVVQKQRAVIRLQKWRSSIRDMRLCKAGYRIRLSLSLQKWR 1440
Db 1381 VQRRYRGNLEMRKQLEVVQKQRAVIRLQKWRSSIRDMRLCKAGYRIRLSLSLQKWR 1440
Qy 1441 ATVOARRQREIFLSTIRKVLMOAFIRATLMLRQORREFEMKRAAVVIQRRFRACAML 1500
Db 1441 ATVOARRQREIFLSTIRKVLMOAFIRATLMLRQORREFEMKRAAVVIQRRFRACAML 1500
Qy 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLQKQFGRKLMTEORNCF 1560
Db 1501 KARQDYQLIQSSVILVQKFRANRSMKQARQEFVOLRTIAVHLQKQFGRKLMTEORNCF 1560
Qy 1561 QLLRCSMPGFQARAGFMARKRFQALMTPPEWMDLIQKRAAKVIQRYWRGYLIRRRQKHQ 1620
Db 1561 QLLRCSMPGFQARAGFMARKRFQALMTPPEWMDLIQKRAAKVIQRYWRGYLIRRRQKHQ 1620
Qy 1621 GLLDIRKRIAGLROBAKAVNSVRCQVQAVRFLGRFTIASDALVLSLDLSLSTVPHL 1680
Db 1621 GLLDIRKRIAGLROBAKAVNSVRCQVQAVRFLGRFTIASDALVLSLDLSLSTVPHL 1680
Qy 1681 MWCSEFMSTFCYGMQAQIRSEVDKOLLERCRIILNLARYNSTNTVNTFOEGGLVTIAQM 1740
Db 1681 MWCSEFMSTFCYGMQAQIRSEVDKOLLERCRIILNLARYNSTNTVNTFOEGGLVTIAQM 1740
Qy 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYVRETKLVARKKMKQ 1800
Db 1741 LLRWCDKDSIFNTLCTLIWVFAHCPKKRKIIHDYMTNPEAIYVRETKLVARKKMKQ 1800
Qy 1801 NARKEPPMTSGRYKSOKINFPTCSIPSLPPDFGIIRSPYTFISSVAFDTILCKLQIDM 1860
Db 1801 NARKEPPMTSGRYKSOKINFPTCSIPSLPPDFGIIRSPYTFISSVAFDTILCKLQIDM 1860
Qy 1861 F 1861
Db 1861 F 1861

```

```

RESULT 3
Q8SX66
ID Q8SX66 PRELIMINARY; PRT; 877 AA.
AC Q8SX66;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE LD39479p.
GN ASP OR CG6875.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STAIN-Berkeley;
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleeb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094825; AAM11178.1; -.
DR FlyBase; FBgn000140; asp.
DR GO; GO:0005875; C:microtubule associated complex; IDA.
DR GO; GO:0008017; F:microtubule binding; IDA.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00307; CH; 1.
DR Pfam; PF00612; IQ; 16.
DR SMART; SM00033; CH; 1.
DR PROSITE; PS00021; CH; 1.
DR PROSITE; PS00096; IQ; 3.
SQ SEQUENCE 877 AA; 106401 MW; 328AC8FD21C027CE CRC64;

```

```

Query Match 44.1%; Score 4200; DB 5; Length 877;
Best Local Similarity 99.4%; Pred. No. 1.9e-265;
Matches 829; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Qy 634 MORTLELLLCFNPWLRLGLEVVFGKIQMSNRDIVGLSTFTILNRLFRNKCEQRYSK 693
Db 1 MORTLELLLCFNPWLRLGLEVVFGKIQMSNRDIVGLSTFTILNRLFRNKCEQRYSK 60
Qy 694 AYTLTEEYAEITKXHSIQKILFLPLDQAKQKRVKHNPCLFVKKSPHKETKDLIRFS 753
Db 61 AYTLTEEYAEITKXHSIQKILFLPLDQAKQKRVKHNPCLFVKKSPHKETKDLIRFS 120
Qy 754 SELLANIGDITRELRRLGYVLOHRTFLDEFDYAFNNLAVLDLGRVLRTRVMEVILLRDD 813
Db 121 SELLANIGDITRELRRLGYVLOHRTFLDEFDYAFNNLAVLDLGRVLRTRVMEVILLRDD 180
Qy 814 LTRQLRVPALSRQRIFNVLKALGALGEANFOLGDDIAAODIVDGHREKTLISLLWOLLYK 873
Db 181 LTRQLRVPALSRQRIFNVLKALGALGEANFOLGDDIAAODIVDGHREKTLISLLWOLLYK 240
Qy 874 FRSPKFAAAATVLOKWRHSHLHVVIQRRIRHKELMRHRAATVIAVFRGHQMRKYVKL 933
Db 241 FRSPKFAAAATVLOKWRHSHLHVVIQRRIRHKELMRHRAATVIAVFRGHQMRKYVKL 300
Qy 934 FKTERTOAAIIILQKFTRRYLAQKQLYOSYHSIIITQWRRAOQLGHQHRQRFVELREAAI 993
Db 301 FKTERTOAAIIILQKFTRRYLAQKQLYOSYHSIIITQWRRAOQLGHQHRQRFVELREAAI 360
Qy 994 FLQRIWRRRLFAKLLAAAEARLQSQKQAAASYIQOWRTYQGRIGRORHEFLRORDL 1053
Db 361 FLQRIWRRRLFAKLLAAAEARLQSQKQAAASYIQOWRTYQGRIGRORHEFLRORDL 420
Qy 1054 IMFVQRMESKWSMLBQRFQOLKKAATINQWRRAKLSMRKCNADYLALRSSVLKQVA 1113

```

Db 421 IMPVORRMRKSWMLRKEFOQKRAAINIQORWRKALSMRKNADYLALRSSVLKVOA 480  
 QY 1114 YRKATTOIMRIDRHHYSLRKNVICLOORLRAIMKMRQORENYILRLNASILVQKRYMRQ 1173  
 Db 481 YRKATTOIMRIDRHHYSLRKNVICLOORLRAIMKMRQORENYILRLNASILVQKRYMRQ 540  
 QY 1174 QMIODRNVLRTRKCIINQVRRWRATLQMRERKNYLHLQTTTKRIQIKFRKREMKKOR 1233  
 Db 541 QMIODRNVLRTRKCIINQVRRWRATLQMRERKNYLHLQTTTKRIQIKFRKREMKKOR 600  
 QY 1234 ABFLQJLKKVTVVQKERRALLQMRKEROBYLHLREVTIKLORRPHAKOQMRFRKRYGT 1293  
 Db 601 ABFLQJLKKVTVVQKERRALLQMRKEROBYLHLREVTIKLORRPHAKOQMRFRKRYGT 660  
 QY 1294 QAAVSCLOHWRNHLRKERNFLOLRQAATLORRYRARNLMIKOLKSYAOLKQAAT 1353  
 Db 661 QAAVSCLOHWRNHLRKERNFLOLRQAATLORRYRARNLMIKOLKSYAOLKQAAT 720  
 QY 1354 IQTRYBAKAMOKOVLYQKORAILIKVORRYRGNLEMRKOIEVYQKORQAVIRLOKQWR 1413  
 Db 721 IQTRYBAKAMOKOVLYQKORAILIKVORRYRGNLEMRKOIEVYQKORQAVIRLOKQWR 780  
 QY 1414 SIRDMLCKAGYRRIRLSLSIQKWRATVQARRQREIFLSTIRKVRMLMQAFIR 1467  
 Db 781 SIRDMLCKAGYRRIRLSLSIQKWRATVQARRQREIFLSTIRKVRMLMQAFIR 834  
 RESULT 4  
 Q8CU27 PRELIMINARY; PRT; 3122 AA.  
 AC Q8CU27;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Abnormal spindle.  
 GN CALMBP1 OR ASP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=Swiss Webster;  
 RX MEDLINE=22242604; PubMed=12355089;  
 RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,  
 RA Askham J.M., Springell K., Mahadevan M., Crow Y.J., Markham A.F.,  
 RA Walsh C.A., Woods C.G.;  
 RT "ASPM is a major determinant of cerebral cortical size."  
 RL Nat. Genet. 32:316-320(2002).  
 DR RMBL; AF533752; AAN46088.1; -.  
 DR MGD; MGI:1334448; Calmbp1.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR000048; IQ\_region.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00612; IQ; 61.  
 DR SMART; SM00033; CH; 2.  
 DR SMART; SM00015; IQ; 55.  
 DR PROSITE; PSS00021; CH; 2.  
 DR PROSITE; PSS00096; IQ; 12.  
 SQ SEQUENCE 3122 AA; 80DED62DE0F4E7B2 CRC64;  
 Query Match 14.4%; Score 1374; DB 11; Length 3122;  
 Best Local Similarity 22.8%; Pred. No. 3.7e-80;  
 Matches 553; Conservative 369; Mismatches 743; Indels 756; Gaps 76;  
 QY 1 MELVWSPVLEVACKETLQILDNRNFRKEVMILKSKNQPKVPKRPFTVGKTLQKSP 60  
 Db 98 ISVTWTLRGGVREIYTVLWN-DFLKQAILL-GNAEPPKKKRSWNNTSK-----KIPA 151  
 QY 61 GAGTKMSV-----VSAAVQOKR-----MSAAAAPPSPQVTRVTPAPS-RPA 101  
 Db 152 SSKHTKTSKNQHFNESFTISQKDIRSPLOPCENLAMSECSSPTEN---KVPTPSIPI 208

QY 102 AWAHPPQAPLV--EKNVYKTPQEEPVYIISOPRSLK-ENLSPMTFGNLLVDIDNLR-F 157  
 Db 209 RECQSETCLPLFRESTAYSSLHE-----SENTQNLKVQDASISQTFDFNEEVANETFIN 263  
 QY 158 PLT---ETGKGQATIPDNLAAMPPTLKGNVKSCAND--MRPRITPDDEDDLED-QPATN 211  
 Db 264 PISVCHQSEGRKLTAPN--CSSPLNSTQTHFLSPDSFVNNRYTSDNDLKMKNVL 321  
 QY 212 KPFVKGHSITNISLTDLD--CSRIDGQHPHLNKTITIVHATHRALACIHEEGSPSP 269  
 Db 322 DTRKDPABSVLCESQTVHEVCQTI--LSPDSFLNDYGLKKGLNFKSV-----NPV 371  
 QY 270 RPTPKSAIHDLRDILKVGSPLEKYSKMLSL-----LSPQTKYAIQSGMPLNEMKI 324  
 Db 372 LSPTQF-----VKDSMGHVQQTGKNEASQDMRINEGLAVTPECQHA---QTPSSRSEKQ 424  
 QY 325 RSIE-----QNRYYQEQQIQIKAKDLNLSSSSSSEASLAGOQEFIPNHS 366  
 Db 425 NPVEVKPHKYDFTKQPKTCFQDAFCHQSKQPKRPFILSATVTKRKTNAREKL---P 481  
 QY 367 EILAQSSRENLIH-----EYGRKSVKG---SPVKNP-----H 394  
 Db 482 EINKPDAKCLEGLVQORKEVGLSREKGFHPSLPVVEFGVSKALSYRDEVTPATVVVAR 541  
 QY 395 KRSHSLSFSDAFSNESLYRNETVAISPCK---QREVDI--TLPRSAAPANASARSSA 449  
 Db 542 KRKSHG-TVGDA--NGKVAEEWMDMCEVKRIHFSPLESTPSTVARTTKKEGLSKRISS 598  
 QY 450 HAWPHAQSKFKLAQTNWGLMKPATPKVRDTSIQPSVKLYDSELYMQTCINPDPAAT 509  
 Db 599 -----LERSGLKKMDSSILKTLPSKTKKRRSI---VAVAQSHL---TFIKPLKAAIPR 647  
 QY 510 TIDPFLASTWYLDQAVDRHQADEFKWLNALVSI PADLDADLN--NKIDVGLKLENEVKNKE 568  
 Db 648 HMPFAKKNFYDERWKEKQEQGFTWNLNLI-F-PDDFTVKTNVSKUNAASVLGASQH 706  
 QY 569 LV---VAPTKEEQSMN--YLTKYRLETLRKAAVELFFSEQMRLPCKSVAVYVNVKQALRIS 624  
 Db 707 KISVPKAPTKEEVSLRAYTASCRLNLRFTACSLFTSEKMKVAKKKEIEIEVGRLLVRK 766  
 QY 625 DRNLHLDVVMQRTILELLLCFNPDLWLRGLLEVVGEEKIQMQSNRDIVGLSTFILNRLFRN 684  
 Db 767 DRHLWKDIGQRKVLNWLSSYNPLWLRIGLETYVFGELIPLADNSDVTGLAMFNLRLWN 826  
 QY 685 K--CESORYSKAVTLTEE-YAETIKKHSLOKILFLPFLDOAKOKRIVKHNPCLFVKKSP 741  
 Db 827 PDIAAEVVRHTVPLPRDGHGAALSKFTLKKLLILCFLDHAKISRLIDHDPCLFCKDAE 886  
 QY 742 HKETKDILLFSSSELLANIGDITRELRRLGYVLQHRQTFLDEFDYAFNNLAVDLRQVRL 801  
 Db 887 FKASKELLALAFSRDLSGEGDLSRHLSTGLPVSQVTPLEDFDFAVNLAVDLQCCVRL 946  
 QY 802 TRVVEVILLRDDLTRQLRVPDAISRLQIFNVKLALGALGANFOL-----GGDIAAQDIVD 857  
 Db 947 VRTVELLTQNNWLSDKLIRIPAIRSVQKMHNVLDVLQVLSRGVPLTDEHGAISSKDVVD 1006  
 QY 858 GHREKTLILLWLIYKFR-----  
 Db 1007 RHREKTLGLLWKTALAFQVDISLNDQLKEEIDFLKHTHSIKRAMSALTCPESQAITNKOR 1066  
 QY 876 -----  
 Db 1067 DKRISGNFERYGDSVOLLMDVWNAVCAYFNKVENFTVSFSDGRVLCYLIIHYHPCYVPF 1126  
 QY 876 -----SKFPH 880  
 Db 1127 DAICORTSQSVACQTSQVVLNSSSEGGCLDLSLEALDHESTPEMYKELLENEKNF 1186  
 QY 881 -----AAATVLOKWR 891  
 Db 1187 LVNSAARDLGGIPAMIIHSDMSNTIPDEKVVIIYVFLCARLLDLRKEIRARLIQTWR 1246  
 QY 892 RHMLHVVIQRRIRH-----KELMRHRAATVIOAVFRGHQMRKY 930



```
Db 1247 KYKL-----KDLKHHQERDKAARVTSQVVNLFSLRRRLQKNVSAALVIQCKWRRVSAQK 1302
Qy 931 VKLEKTER-----TQAAIILQKFTRYLAQKQLYOSYH-----963
Db 1303 LRLMLKNEKLAKONKSAALQIYWRYSYTRKFRFLKHYSTVLOSIRIMKIALTSYKRYL 1362
Qy 964 -SIITIORWRAQQGROHROFVELREAAFLQ---RIWRRRLFAKLALAAETARLQR 1019
Db 1363 WATVTIQRHWAYLSGKRDQIIFRLKSSSLVQIFWFRWRKRLQLOTKAAVTLQRAFR 1422
Qy 1020 S---QKQ-----QAAASY-----1029
Db 1423 EWHLEKQIRERSAVVIOQWYRMHRELQKVIYIRSCVIVIORVRFCQAKLYKRRKDAIL 1482
Qy 1030 -IQOWWRTYQLGRIQHEHFLORDLIMFVOREMR-----SKWSMLE 1069
Db 1483 TLQHYRARQKGLAHADYDILQRAATIRLOAARFGMKARHSYRLQIGAACVLQSWRRWQ 1542
Qy 1070 QR-----KFPQOL-----KRAAINIQORWRKLSMRKCNADYLALRS 1106
Db 1543 ERVRFLNLKMWIKLQAHIRKYQOLQYKVKIKKAAITITQHFASISARRVLASQYKTRS 1602
Qy 1107 SVLKVQ-----AYRKATIQWRIDENHY--YSLRK---NVLCQORLRAIMKREQ 1151
Db 1603 SVTLQSAACRGWQARKAFRHALASVIKIQSYRYAYICRKTQNFKNATIKLSIVMKQS 1662
Qy 1152 RENYLRLRNASILQKRYMRQOMIQDRNAVILRTKCIINVOORWRATL-----1200
Db 1663 RKQVLOIRAAALFIORWTRSQKLASQKKEVIQVRESCKIQSHPRGCLVRKQLKQCKA 1722
Qy 1201 -----OMRRERNYHLQTTTKRIQIKPRAKEMKKQRAEFLQLKKVTLVVO-----1247
Db 1723 AISLOSYPERMRTARQYILKMKCAALVIOSFYCAVEAQISQRKNFQVQKRAAICLQAAVRG 1782
Qy 1248 -KRRALLQOM-----RKEROYHLREVTIKLORRHFQAKSMRPMKAYRG 1293
Db 1783 CKVRQIQKQSTAAVTIORVFRGHSQRMKYQTMLOQSAVKIORWRAQKVAAYDMRIFOQLKT 1842
Qy 1294 QAAVSCLOMWRN-----HLLR-----1310
Db 1843 REAVVCLQSAVSGWQVROOLRRQHEAAVVKIOSTFRMAVAQQOYKLLRAAAVIOQHVRAR 1902
Qy 1311 ---KRRNSFLQROAAITLQRRVRARL-----NMIKOLKSYAQLK 1348
Db 1903 AAGKRQHLAYTLQRLHAALVFOAAWKGWMLRQIARQHQCAALIOSYYRMHIORRKSIMK 1962
Qy 1349 QAAITIQTRVRAKAMQOVLYQKQREAILIKVQERYEGNLEMRKQI-----1395
Db 1963 TAALQIQLCYRAYKVGKEORHLYLTKAAAVVTLQSAVRG-MKVRKRVAECHKAAVTTQSK 2021
Qy 1396 -----EYQKQOQAVIRLOKWRRSIR-----1416
Db 2022 FRAYRTQKTYTIRTSALVIOQWYRNKITTQOQHOEYLNLRRAAVOQAAVYRGITVRRI 2081
Qy 1417 -----DMRLCKAGYRRIRLSSUSIQKWRATVOARQORRIFLSTIRKVRML 1462
Db 2082 QHMHMAATLIEAMPKMRQSRVYLLKMRATAALIIQVRVAYVYLGKIQHEKYLRTLKAITL 2141
Qy 1463 QAFIR-----ATLLM-----ROOREEFEMKREAAVVIQRRFEARCAMLKAR 1503
Db 2142 QAGVGRARVTRVKMHFAATLIIQSHFRGHQQTTFHRLKAAATVVOORYFAVKEGSAEF 2201
Qy 1504 QDYQLIQSSVILQKFRANRSMQ-----ARQEFVQLRTIAVHLQ 1544
Db 2202 QRYSLRRESVLLIQAAFRGLTRRHLKAMHLAATLIQRRFRTFAWRKFLSLRKTALWIQ 2261
Qy 1545 QKFRGKRLMIQORNCQFOLLRCMFGFOARAGFMWRKRFQAL-----MTPEMM 1592
Db 2262 ROYRA-RLYAKYSRQOLLLEKAVTIKIOSSVRGVVRKRVQKQHPAAATVIOATFRMHGAYM 2320
Qy 1593 DLIRQKAAKVIQRYWRGYLIRRRQKHQGLDLDIRKRIALQRLQREKAVNSVRCKVQEAVR 1651
```

```
Db 2321 RYQHLKRASVVIQVHTAAEL--QROKHAAVILOAAVRGMKTRSHLKTWSSATLIQSOPR 2378
Qy 1652 --FLRGRTIASDALAVLSQDLSELTVPHELLMWCSEFMSTFCYGTMAQAI--RSEVDKOL 1707
Db 2379 AFIVRRRRIARLKAIFAQVQ--RKFRATLYAKHJHOFLOLRKAAITIOSSVYRLMVQK 2436
Qy 1708 IE--RCSRIILNLARYNSTTV 1726
Db 2437 QEMHRAALIQAATFEMHRTYV 2457

RESULT 5
Q81ZT6
ID Q81ZT6 PRELIMINARY; PRT; 3477 AA.
AC Q81ZT6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Abnormal spindles.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Ashkam J.M., Springall K., Mahadevan M., Crow Y.J., Markham A.P.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AF509326; AAN40011.1; -.
DR Genbank; HGNC:19048; ASPM.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00307; IQ; 2.
DR Pfam; PF00612; IQ; 71.
DR SMART; SM00033; CH; 2.
DR SMART; SM00015; IQ; 63.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS00096; IQ; 10.
SQ SEQUENCE 3477 AA; 409796 MW; 49AE76DC75124980 CRC64;

Query Match 14.3%; Score 1360; DB 4; Length 3477;
Best Local Similarity 22.7%; Pred. No. 3.6e-79;
Matches 480; Conservative 360; Mismatches 630; Indels 640; Gaps 68;

Qy 5 WSPVLEVACKETLQIDNFRKVMILKSKSNQPVKNPRKFPPTVGKTLQKSGTGAGK 64
Db 110 WTPLEGRVREIMTFLVN-DVLKHQAILLQNAEQQKKKRLWDTI-----154
Qy 65 TMKSVSAAVQOQKRRMS-----AAAAPPKQTMVATAPSRPAWAHPPQAPLYEKVYK 119
Db 155 -KKKKISASTSHNRKRVSNIQNVNKTFSVQKVRSLPQACENLAMNEGPPPTENSL- 212
Qy 120 TPQEPYVISP-----QPRSLK-----ENLSPMTPGNLLDV-----IDNLRFT 157
Db 213 ILEENKIPISLPAFNECHGATCLPLSVRRSTYSILHASENRELNLNVHSANVSKVSFN 272
Qy 158 --PLTET-----RGKQATIFPDNLAAMPPTPKLGNVKSCAN--DMRPRRTIPDD 203
Db 273 EKAVTETSFNVSNNVNGQGENSKL-----SLTPNCSSTLNTOSQIHFLSPDS 320
Qy 204 LEDQPATNKTFDVKHSETINISLDTLDCSRID---GQPHTPLNKTTTIVHATHRALACI 260
Db 321 P-----VNSHGANNELVTVCLSDMPKNSQPVHLESTIAHVIYKIL---366
Qy 261 HEEGPGPPPTPKSAIHD---LKRDIKLYG-SPLRKYSESMDLSLLSPQTKYAIQSGM 316
Db 367 -----SP-----DSFIKDNVGLNQLSESVNFIPLSPNOFLKDNMAYMCTSQCTCKVPL 415
Qy 317 PNLNEMKIRSIQRYQYQEQIQIKAKDLNSSSSSEASLACQOQFPLNHSILQASS-RF 375
```



Db 416 SNEQSVQSPED---WRKSEVSPRIPECGSKSPKAIPEELVEMKSNYYSPFKQNNPKF 472  
 QY 376 N-LHEVGRKSVKSPKNP---HKRS-----HELSP 403  
 Db 473 SAVQDISSHHNKPQRRPILSATVTKATCTRENQTEINKPKAKRCLNSAVGEHKVI 532  
 QY 404 SPAPSNESE-----YRNE-----TVAISPPKKQ--RVEDTTLPRSAAP 439  
 Db 533 NNQKEKEDPHSVLPIDPILSKSKSYKNEVTPSSITASVARKEKSDGSMEDANV-RVAIT 591  
 QY 440 ANASARS-SAAIAWHAQ-----SKFKLAQWMSLMKP-----ATPRKV 478  
 Db 592 EHTVEVREIKRIHFSPSEPKTSVAKTKNVTTPISKRSINREKLNKKTKDLSIFRTPISK 651  
 QY 479 RUTSIPQSVKLYDSELYMOTCINPDFAATTID-----PFLASTMYLDQAVDRHQADF 534  
 Db 652 TNKRTKPIITAVAQSSL---TFIKP-----LKTDIRHPMPFAAKMFPYDERWEKEQOGFT 704  
 QY 535 KWLNALVSIIPADLDADLN-NKIDVGKLFNEVRNKLWV---APTKEEQMN-YLTKYRLE 589  
 Db 705 WMLNFLT-PDDFTVKTNISVNNAAATLLGIENQHKSIVPRAFTKEMSLRAYTARCLN 763  
 QY 590 TURKAAVELFBEOMELPCSKVAVYNNKQALRTSRDNLHLDVVMORTLELLLCNPLW 649  
 Db 764 RLRRACRLFTSEKMVKAIKLEIETEARLIVRKDRHLWKDVGQRQKVLNMLLSYNPLW 823  
 QY 650 LRIGLEVFGEKIQMOSNRDVLGLSTFIILNRLFN---KCEQORYSKAYTLTEEYAEATIK 706  
 Db 824 LRIGLETTVGLLSLSDNSDVTGLAMFILNRLNLPDIAAEVYHPTVPHLYRDGHEALS 883  
 QY 707 KHSLOKILFLPDLQAKQKRIWKHNPCLFVKKSPHETKDILLRFPSELLANIGDITRE 766  
 Db 884 KFTLKKLLLVCLFDYAKISLIDHDPCLFCKDAEPKASKELLAPSRDFLSEGDLSRH 943  
 QY 767 LRRLGVVLQHQFTLDEFPYAFNNLAVDRDGVRLTRVVEVILLRDLRQURVPAISRL 826  
 Db 944 LGLLGLFVNHVQTFDFDFAVTNLAVDLQCGVRLVKTWELLTQNWDLSSKLRIPASRL 1003  
 QY 827 QRIENVKLAGALGEANFOL---GGDIAAQDIDVGHREKTLISLWLIYKPR-----875  
 Db 1004 QKMHNVDIQVLQVLSKRSIELSDHGNLTLSKDIVDRHREKTLRLWKIAPAFQVDLSML 1063  
 QY 876 -----875  
 Db 1064 DQKKEBIAFLKHTSIKKTISLSCHSDDLINKKKGKRDSPGSEQYSENIKULMDWNAV 1123  
 QY 876 -----875  
 Db 1124 CAFYNNKVENFTVFSRGRVLCYLIHHYHPCYVFPDAICQRTQTVECTQTSVNLSSS 1183  
 QY 876 -----SPKFH-----880  
 Db 1184 ESDSDSLMSLKAFDHENTSSELYKELLEKNFHLVRSNVRDLGGIGPAMINHSDMSNTI 1243  
 QY 881 -----AAATVLQKWRHRLHVLVVIQRRIRHKELMRHRAAT 916  
 Db 1244 PDEKVVITVLSFLCARLLDLRKEIRAARLIQTWRKYKLTDLK---RHOE---REKAAR 1297  
 QY 917 VQAVFRGHQMRKYVKLFTERTQAAIIQKFTFRVLAQKOLYQ-----960  
 Db 1298 IIQLAVINFLAKQRLR---KRVNAALVIQKYWRVLAQRKLLMLKKEKLEKQVONKAASL 1353  
 QY 961 -----SYHSII-----TIQWRWRAQQLGROH 981  
 Db 1354 IQQWRYRSTQRFPLKLYSIILOSIRIMIIVATYSKRYLWATVTIQRHWRAYLRKOD 1413  
 QY 982 RQRFVBLREAAIFLQ---RIWRRLPFAKLLAAETAAR-----LORSQKQAAASVIOH 1032  
 Db 1414 QQQYEMLKSSSTLIIQSMFRKQKQKQMSQVATVILQRAFREWHLRKQAKENSAILIOS 1473  
 QY 1033 QWRTYQIGRIQREHFLQRDLIMFVQRRMRKSWMLERQKBFQOLKRAAINIQORWRAKL 1092  
 Db 1474 WYRMHK-----ELRKYTIYRSCVVIQKRFR---CFQAQKLYKRRKESILTIQKYKAYL 1525

QY 1093 SMKCNADYLALRSSVLKVQA-----YRKATI-----OMRIDRHNHYSLRKN 1134  
 Db 1526 KGIERTNLQKRAAAIQOAAFRRLKAHNLQRIIRAACVIOSYWNRQDRVFLMLKKT 1585  
 QY 1135 VTCLOQRLEAIMKMBQRENYLRLNASTLVOKRYEMRQOMIQDRNAYLRTKRCIINVOR 1194  
 Db 1586 II---KFOAHRVKHQORQKYKMKKAAVIOHPRAYIFAMKVLASYQKTRSAVIVLOS 1641  
 QY 1195 RWRATLQMRERKNVYLHOTTTKRTOIKFRA---KRE-----MKQRAE 1235  
 Db 1642 AYRG-----MQARKMTHILTSVIKQSYRYRAYYSKEFLSKNATIKLOSITVKMKQTRKQ 1697  
 QY 1236 ELQGLKVTLVVOKRRALLQMRKEROEYHLHRAVTIKLO---RRFHAQSMRPMRAKYRG 1292  
 Db 1698 YLHLRAALFIQOCTRSKKIAQKREYEMQMBESCILQAFVRYGLVRKOMRLQ---1752  
 QY 1293 TOAAVSCLOMHRNHLRLKRERNSTFLOLRQAATTLQRRYRARNLMIKOLKSYAQLKQAAI 1352  
 Db 1753 --KAVISLQSYFR---MRKARQYLYKMYKAILIIVQNYHAYKAVQNRKNFLQVKKAAAT 1806  
 QY 1353 TITQTRYAKKAKQKVVLVYKOBELIIVQRRYRGHLEMRKQLEVYQKORQAVIRLOKWW 1412  
 Db 1807 CLQAAVRYGYKVRQ---LIKQOSIAALKIQSAFRG---YNNKRVK-YQSVLOSIIKIQRWY 1858  
 QY 1413 RSIRDMLCKAGYRIRLSSLSIQ---RWRATVQARROREITFLSTIRKVLWMAFIRAT 1469  
 Db 1859 RAYTLHDTRTFLTKAAVISLOSAYRGWKVKQIRREHO-----1899  
 QY 1470 LLMRQORREFEMKRRAAVVIQRRFRARCAMLKARQDYQLIOSSVILVQRKFRANRSMKQA 1529  
 Db 1900 -----AALKIQSAFR---MAKAQKQFRLFKTAALVIQNFRAWTAGRKQ 1940  
 QY 1530 ROEFVOLRTIAVHLQOKFRGK---RLMIEQRNCFQLLRCSMPGFOBARAGFMARKPEQAL 1586  
 Db 1941 CMYEIHELRAVLVLOSMMWKGKTLRLQOLQOHKCAII-----QSYRYRMHVQOKKWKIM 1993  
 QY 1587 MTPEMMDLIRQKRAAKVIOQRYWRGYLIRRRQKHGL-----LDIRKRIAQ 1631  
 Db 1994 -----KQALLIOKYTRAYSIGREONHLYLKTAAVVTLOSAYRGMKVRKRIKD 2042  
 QY 1632 LRQEAKAVNS 1641  
 Db 2043 CNKAAVTIQS 2052  
 RESULT 6  
 QSN4D1  
 ID QSN4D1 PRELIMINARY; PRT; 1142 AA.  
 AC QSN4D1;  
 DT 01-OCT-2002 (TREMELrel. 22, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Similar to asp gene product.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC034607; AAH34607.1; -.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR001715; Calponin-like.  
 DR InterPro; IPR000048; IQ\_region.  
 DR Pfam; PF00307; CH; 2.  
 DR Pfam; PF00612; IQ; 14.  
 DR SMART; SM00033; CH; 1.  
 DR SMART; SM00015; IQ; 11.  
 DR PROSITE; PS50021; CH; 2.  
 DR PROSITE; PS50096; IQ; 2.

SQ SEQUENCE 1142 AA; 133617 MW; FA5738B59F20B83A CRC64;

Query Match  
Best Local Similarity 9.1%; Score 868; DB 4; Length 1142;  
Matches 336; Conservative 217; Mismatches 492; Indels 322; Gaps 46;

Qy 582 YLTYYRIETLRKAARELVFFSQBELPCSKVAVYVYNNQALRIRSRDNLHLVDVVMQRTIEL 641  
Db 6 YTARCLNLRRAACRLFTSEKWKAKLLEIEARPLIVRKDRHLWKDVGEBQKVLNW 65

Qy 642 LLCNPNLWRLGLEVFGKEIKOMQNSRDIVGLSTFILNRLPN---KCEORYSKAYTLT 698  
Db 66 LLSYNPLMLRTGLETTYGELISLSDNSDVTGLAMFILNLLWNPDIAAEYRHPVPHLYR 125

Qy 699 BEYAEETIKKGSLOKILFLPLDOAKOKRIIVKHNPLCFVKKSPHKEWDIILLRESSLLA 758  
Db 126 DGHBEALSKFTLKKLLLVLCVDFVAKTSLRDHPCLPCKDAEFKASKBELLASRDFLS 185

Qy 759 NIGDITBELRLGVLQHQRTFLDEFDAFNLAVALDRDGVRLTRVVEVILLRDLTRQL 818  
Db 186 GEGDLSRHLLGLGLPVNHVQTPFDEFDAVTLNLAVALDQCGVRLVTRMELLQNWDLSSKL 245

Qy 819 RVPALSRLORFNVKLAIGALGEANFQI---GGDIAAQDIDVGHREKTLISLLWLIYKF 874  
Db 246 RIPALSRLOKMHNDIIVLQVLSRGIELSDEHGNITLSKDIVDRHREKTELLWKIAPAF 305

Qy 875 RSPKFHAAATVLQKWRHRLHVLVQRIIRHKLMRHRAATVIAQVPRGH--QMRKYVK 932  
Db 306 Q-----VDISINDLQKE-----EIAFLKTKTSIKKTIS 334

Qy 933 LFKTERTQAAILQKFTTRYLAQKQYSHSIITIQWRMA--QOLGRQHRQFVELRE 990  
Db 335 LLSCHFD--LINKKKGKRDGSGSEQYS--ENIKLLMDVNAVCAFYNNKVENFTVSPSD 390

Qy 991 AAIF-----LQRTWRRLFAKLLAAETHARLQRSOKQOQAAAQYIQWQRTY 1037  
Db 391 GRVLCYLIIHHYHPCYFPDAICQ---TTQVECTQTGSSVVLNSESDDSLDMSLKAF 447

Qy 1038 QLGRIQRHEFLRDLIMFVQRRMSKWSMLEQKEFQOLKRAAINIQORWRKLSMRKC 1097  
Db 448 D-----HE--NTSELYK-----ELLENKKNEHLVRSVR----- 475

Qy 1098 NADYALRSSVLKQAVKATQIMRIDRNHYSLRKNVICLQRLRAIMKWRQRENYLR 1157  
Db 476 --DLGGIPAMI-----NH--SDMSNTI-----PDEKVV 499

Qy 1158 LRNASILVQKRYRQOMIQDRNAYLTRKCIINVQRRWATLQMRERKNYLHLQYTK 1217  
Db 500 ITYLSFLCARLLDLEKEI---RAARL-----IQTTWR--KYKLTDLKQHBREKAA 547

Qy 1218 RIQ---IKFRAKREMKQRAEFQLKKVTLVVQK-----RRRALLQMRKQRYLHLRE 1268  
Db 548 IIQAVINFLAKQRLRK-----VNAALVIQKWRVLAQRLMLLKKLEKLVQNK 600

Qy 1269 VTI--KLQRRPHAQKSMRPMKRYGTGTAQVSCLOMWRNH-----LLR 1310  
Db 601 ASLIQAMWRRYRAKYLCKVKAACK--IQAWYRC---WRAHKEVLAALKVAKLIQGCFT 655

Qy 1311 KEENSFLQROAATLTQRYRABRLNMIKQKSAQLK--QAAITQTRYAKYAMQOV 1368  
Db 656 KLETRFLNVRASALITQKWRALIPAKIAHEHFLMIKRHAACLIQAHYGYKG--RQV 713

Qy 1369 VLYQKQRAILK--VQRRYRGNLMRKQIEVYQORQAVIRLQKWRISIRDMRLCKAGYR 1426  
Db 714 FLRQSAALIIQKVIAREAGHERIKYIEF-----KKSIVLQ----- 752

Qy 1427 RIRLSLSIQKRWATVQARQREIFLSTIKRIMQAFTRATLLMRQORREFEMKRAA 1486  
Db 753 -----ALVRGLV-----RKFLEQRAKIRLLHFTAAAYHLN-----A 786

Qy 1487 VVIQRRFRACMLKARDYQLIQSSVILVQKREFRANESMKAQOEYVQLRTIAVHLQX 1546  
Db 787 VRIQRAYKLYLAVKANKQV-----NSVICIQWRFAQLQEKRFIQKTHSI----- 832

Qy 1547 FRGRLMIEQRNCFOLLRCSPGFOARARGFMARKRFQALMTPEWMDLIROKRAKV--I 1604  
Db 833 ---KKIEHGEQECLSQRNRAASVIOKAVRHLLRK-----QKFTSGIITKI 876

Qy 1605 QRYWRGYLIIRRRKHQGLLDIRKRIAQLQROPAKAVNSVRCVQBAVRFGRFIASDALA 1664  
Db 877 QALWGGYSWRKNDCTKIKALRSLQVNVNREIBENKLYKRTALAHYLLTYKHLSAILE 936

Qy 1665 VLSOLDRLSRTVPHLLMWCSEFMSTFCYGINAQAIRSEVDQOLIERCSR----- 1713  
Db 937 ALKHELVVTR-----LSPLCCENNAQSGAISKIPVLIRSCNRSRIPCMEVIRYA 984

Qy 1714 --IILNLARVNSTTNTVFOEGGLVTIAQMLL-----RWCKDKDSIFNTLCTLIWVF 1762  
Db 985 VQVLLNVSKYEKTTSAVDVENCIDILLELLOIYREKPGNKVADKGSIFTKTCLLAIL 1044

Qy 1763 AHCPKKRKIIHDYMTNPEAIYVMRETKLVARKMKMKONARKPPMTSGRYKSQK----- 1817  
Db 1045 L---KTNRASDVRSRQVVDRIYSLYKLTAKHKM--NTERI-----LYKOKKNSSIS 1093

Qy 1818 INFWPCS-----LPSLEPDPFGIIRYSPYTFISSVYAFDTILCKLQI 1858  
Db 1094 IPFIPETPVTRIVSRUKPDWLRDRDNWEEITNPLQAIQWVMDTLGI 1140

RESULT 7  
Q8IU2 PRELIMINARY; PRT; 825 AA.

AC Q8IU2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Abnormal spindle protein ASP (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon adenocarcinoma grade II;  
RX MEDLINE=2242604; PubMed=12355089;  
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,  
RA Askham J.N., Springall K., Mahadevan M., Crow Y.J., Markham A.F.,  
RA Walsh C.A., Woods C.G.;  
RT "ASPM is a major determinant of cerebral cortical size.";  
RL Nat. Genet. 32:316-320(2002).  
DR EMBL; AY099893; AAM44122.1; -;  
DR InterPro; IPR001715; Calponin-like.  
DR Pfam; PF00307; CH; 2.  
DR PROSITE; PS50021; CH; 2.  
FT NON\_TER 1  
FT TER 825  
SQ SEQUENCE 825 AA; 94804 MW; EED55486C4C88BBB CRC64;

Query Match 7.7%; Score 733.5; DB 4; Length 825;  
Best Local Similarity 34.1%; Pred. No. 3.6e-39;  
Matches 198; Conservative 104; Mismatches 210; Indels 69; Gaps 18;

Qy 334 QEQIQI---KAKD-LNSSSSSEASLAGQOQFLFNHSE-----IIAQSRF 375  
Db 100 RENQTEINKPKAKCLNSA-----VGEHEKVINNQKEKEDPHSYPIIDPILSKSKSY 152

Qy 376 NLHEVGKSVKSGPVKNPHKRSHELFSAPSNEISLYNETVAISPPKQVEDTTLPR 435  
Db 153 K-NEVTSSTTASVA---RKRKSDGSMEDA-----NVRVAITEHTEVREIKRIHFS 199

Qy 436 SAAPANASARSSSAHAPHQAQSKFKLAQTWLSMKXP-----ATPRKVRDTSIQPSVKLY 490  
Db 200 PSEKTSVAKKTKNVTTP--ISKRLSNREKLNKKKTDLSIFRTPISKNRTKPIIAPA 257

Qy 491 DSELYMOTCINPDPFAATTITD---PFLASTWYLDQAVDRHQADEKKLNALNISPAD 546

```

Db 258 QSSL---TFIKP-----LKTDPHPMPFAAKMFDYDERWKEQEQGFTWWLNFILT-PDD 309
QY 547 LDADLN-NKIDVCKLFNEVENKELV---APTKEQSMN-YLTKYRLETLRKADELFFS 601
Db 310 FTVKTNISEVNAATLLGIENQHKISVPRAPTKEMSLRATYARCLNRLRRAACRLFFS 369
QY 602 EQMRLPCSKVAVVNVKQALRISDRNLHLDVVMQRTILELLCFNPLWLRGLGVVFEK 661
Db 370 EKMVKAKKLEIEIARRLLVRKDRHLWKDVGQRQVNLWLLSYNPLWLRIGLLETYYGEL 429
QY 662 IQMOSNRDVLGLSTFILNRLFRN---KCEBQRYSKAYTLTEEYAEITKHSLOKILFLLP 718
Db 430 ISLEDSDVTGLAMFILNRLWNPDIAAEYRHPVPHLYRDGHEEALSKEFTLKLILLVC 489
QY 719 FLDQAKQKIVKHNPCLFVKKSPHKETKOILLRFSSELLANIGDITRELRRLGVVLOHRQ 778
Db 490 FLDYAKISRLIDHDPCLFCDAEFKASKEILLAFSRDFLSGEGDLRSHLGLGLPVNVHVQ 549
QY 779 TFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPAISRLQRIFNVKLAGA 838
Db 550 TPDDEFDAVNLAVDLQCGVRLVTRMELLTQWDLSSKLRIPAIISRLQXMHNVDIVLQV 609
QY 839 LGEANFQ-----GGDIAAQDIVDGHREKTLISLWLIYKFR 875
Db 610 LKSRGIELSDEHGNTILSKDIVDRHREKTLRLWKTAFAPQ 650

RESULT 8
Q81ZJ8 PRELIMINARY; PRT; 826 AA.
AC Q81ZJ8;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Abnormal spindle protein ASP (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Askham J.M., Springell K., Mahadevan M., Crow Y.J., Markham A.F.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AY099891; AAM44120.1; -.
DR InterPro; IPR001715; Calponin-like.
DR Pfam; PF00307; CH; 2.
DR PROSITE; PS50021; CH; 2.
FT NON_TER 1
FT NON_TER 826
FT SEQUENCE 826 AA; 94875 MW; 970ED55486C4C88B CRC64;

Query Match
Best Local Similarity 34.1%; Pred. No. 3.6e-39; Length 826;
Matches 198; Conservative 104; Mismatches 210; Indels 69; Gaps 18;

QY 334 QBQIQI---KAKD-LNSSSSSSASLAGQOEFNFHSE-----ILAQSSRF 375
Db 100 RENQTEINKPKACRLNSA-----VGEHKYNNQKEKEDPHSYLPIDPILSKSKY 152
QY 376 NLHEVRKSVKSPVKNPKHRSHELFSFSDAPSNESLYNETVAISPPKKQKQVEDTTLPR 435
Db 153 K-NEVTPSSTTASV---RKRKSDGSMEDA-----NVRVAITEHTVEIRKIHFS 199
QY 436 SAAPANASARSASAHAWPHAQSKKFLAQTMSLMKEP-----ATPRKVDSTQPSVKLY 490
Db 200 PSEPKTSVAKKTKNVWTP--ISKRIISNREKLNKKKTDLSIFRTSKTKTKPIIAYA 257
QY 491 DSELYMQTCINFPFAATTTID-----PFLASTMYLDEQVDRHQADFKKWLNALVSIAD 546

```

```

Db 258 QSSL---TFIKP-----LKTDPHPMPFAAKMFDYDERWKEQEQGFTWWLNFILT-PDD 309
QY 547 LDADLN-NKIDVCKLFNEVENKELV---APTKEQSMN-YLTKYRLETLRKADELFFS 601
Db 310 FTVKTNISEVNAATLLGIENQHKISVPRAPTKEMSLRATYARCLNRLRRAACRLFFS 369
QY 602 EQMRLPCSKVAVVNVKQALRISDRNLHLDVVMQRTILELLCFNPLWLRGLGVVFEK 661
Db 370 EKMVKAKKLEIEIARRLLVRKDRHLWKDVGQRQVNLWLLSYNPLWLRIGLLETYYGEL 429
QY 662 IQMOSNRDVLGLSTFILNRLFRN---KCEBQRYSKAYTLTEEYAEITKHSLOKILFLLP 718
Db 430 ISLEDSDVTGLAMFILNRLWNPDIAAEYRHPVPHLYRDGHEEALSKEFTLKLILLVC 489
QY 719 FLDQAKQKIVKHNPCLFVKKSPHKETKOILLRFSSELLANIGDITRELRRLGVVLOHRQ 778
Db 490 FLDYAKISRLIDHDPCLFCDAEFKASKEILLAFSRDFLSGEGDLRSHLGLGLPVNVHVQ 549
QY 779 TFLDEFDYAFNNLAVDLRGVRLTRVVEVILLRDDLTRQLRVPAISRLQRIFNVKLAGA 838
Db 550 TPDDEFDAVNLAVDLQCGVRLVTRMELLTQWDLSSKLRIPAIISRLQXMHNVDIVLQV 609
QY 839 LGEANFQ-----GGDIAAQDIVDGHREKTLISLWLIYKFR 875
Db 610 LKSRGIELSDEHGNTILSKDIVDRHREKTLRLWKTAFAPQ 650

RESULT 9
Q81ZJ9 PRELIMINARY; PRT; 882 AA.
AC Q81ZJ9;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Abnormal spindle protein ASP (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retal brain;
RX MEDLINE=2242604; PubMed=12355089;
RA Bond J., Roberts E., Mochida G.H., Hampshire D.J., Scott S.,
RA Askham J.M., Springell K., Mahadevan M., Crow Y.J., Markham A.F.,
RA Walsh C.A., Woods C.G.;
RT "ASPM is a major determinant of cerebral cortical size.";
RL Nat. Genet. 32:316-320(2002).
DR EMBL; AY099890; AAM44119.1; -.
DR InterPro; IPR000048; IQ region.
DR Pfam; PF00612; IQ; 34.
DR SMART; SM00015; IQ; 33.
DR PROSITE; PS50096; IQ; 2.
FT NON_TER 1
FT NON_TER 882
FT SEQUENCE 882 AA; 107102 MW; 84AD920618B6A1F6 CRC64;

Query Match
Best Local Similarity 7.2%; Score 688; DB 4; Length 882;
Matches 225; Conservative 173; Mismatches 314; Indels 206; Gaps 28;

QY 868 WLIYKFRSPKFAATVLOKWRH-----WLH-----VVTQRRIR-----HKELMRRHR 913
Db 6 WHL---RQAKEENSAIIQISWYRMHKLRYIRSCVVIQKFRFCFQAKLYKRRKE 62
QY 914 AATVIAQVPRGHQMKYVKLFKTRTQAIILIQKTRVYLAQKQYQSHSIIITQWRWR 973
Db 63 SILTIQYKAYLKGKIERNTYLRQKRAAIQQAFFR--LKAHNLGRQTRACVIOQYWR 121
QY 974 AQQ-----LGRQHRQRFVELREAAIFLQRIWRRRLFAKLLAAAF 1014
Db 122 MEQDRVRFNLKTTIKFOAHVYRKHQKQYKMKKAAVIIQTHFRAYIFAMKVLASYQK 181

```

1015 AR-----LORSQK-QAAAAAYIQ-----MOWRTYOLGRIOHREFLRDLIMFVORMRS 1063  
 182 TRGAVIVQSAIRGQWARKMVIHILTSVIKIQSYRAYVSKFELSKNATI-----KLOS 237  
 1064 KWSMLEQKEFQQLKRAAINIQWRRAKLSMRKCNADYLALRSSVLKQVAY-----1114  
 238 TVKMKQTRKQYLHLRAALFLIQCYRSKIIAAQKREBEMQWRESCIKLQAFVGRVLRKQ 297  
 1115 -----RKATI-----QMRIDNHYSLKKNVICLQORLRAIMKREBENYLRNLSIL 1164  
 298 MRLQKAVISLQSYFRMRKARQYLLKMYKAILIIVYTHAYKAQVNRKNFLQVKAATC 357  
 1165 VQKRYR-----MRQQMI-----1178  
 358 LQAYRGYKVPQLIKQSIALKTSQAPRGYNKRVKQSVLQSIILKIQWRYAYKTLHDT 417  
 1179 RNAYLTRKCIINVQ-----RRWRATLQMRERKNYHLQTTTKRIQIKPRAKREMKQRAE 1235  
 418 RTHFKTKAAVTSIQSAYRGWKVKQIRREHQAAL-----KIQSAFR-----MAKAQKQ 466  
 1236 FLQJLKVTLVQKRRALLQMKERQEVHLHREVTIKLO-----RRPHAQKS-----1282  
 467 FLFKTAALVIOQNFRAWTAGRKQCMYEIHLHRAVLQSMWKGKTLRROLQROHKCAII 526  
 1283 -----MRFBKAYRGTOAAVSCLOMHWNRHLKRENSFLQRLQAAITLQRRYARLN 1336  
 527 IQSYRMHVQKKNKIMKKAALLIQKYRAYSIGEQNHLYLTKAAVVILOSAYRG-----583  
 1337 MIKOLKSYAQJQAAITQIOTRYAKKAMQKVLYQKOREAIKQVRRYRGNLEMRKQIE 1396  
 584 -MKVRKRIKCNKAAVTIQSYRAYKTKK-----YATYRASAILIQWYRGKITNEHQH 638  
 1397 VYQKQRAVIRLOKWRISIR-----DMELCKAGYRRIRLSLSLSTOR 1437  
 639 EYLLKTKTAIKIQSYRGIRVERHIOHMRATFIKAMFKHQGRISYHTMKRAAIVIQV 698  
 1438 KWRATVQARROREIFLSTIRKVLMOAIFRATLMLQORREFEMKRRAAVVIQRRFRARC 1497  
 699 RCRAVYQKQKREKYLITLAKAVILOASFRGV-----RVRRILKMQVTAATLIQSNYR---751  
 1498 AMLKARQD--YOLIQSSVILVORFPRANSWKQARQEVQIURTIAVHLQKFRGKRLMIE 1555  
 752 ---RYRQOTYFNKLKTKITVQORYWAMKERNIQRYNKLRSYIVYQIAIFRGKK-----804  
 1556 QBNCFQLLRCSMPGFOARARGMARKRFOALMTPEMDLIRQKAAKVIQRYWGRYLIR 1615  
 805 ARRHLMWHIAATLIQRRFRILMMRRRFLSL-----KKTAILIQRYRAHLCTK 853  
 1616 RQKHQGLDIRKRIAQLR 1633  
 854 H--HLQFLQVQNAVIKIQ 869

RESULT 10

Q9NVS1 ID Q9NVS1 PRELIMINARY; PRT; 726 AA.  
 AC Q9NVS1  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ10549.  
 OS Homo sapiens (Human)  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.

RT "NEDO human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK001411; BAA91676.1; --  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR000048; IQ\_region.  
 DR Pfam; PF00612; IQ; 15.  
 DR SMART; SM00015; IQ; 9.  
 DR PROSITE; PS00096; IQ; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 726 AA; 85590 MW; BDC1D2557E7BD244 CRC64;  
 Query Match 4.3%; Score 408.5; DB 4; Length 726;  
 Best Local Similarity 20.9%; Pred. No. 5e-18;  
 Matches 202; Conservative 147; Mismatches 321; Indels 297; Gaps 32;  
 QY 936 TERTQAAIILQKFTRRYLAQKOLYSYHSI-ITIORWRAQQLGRQHRQFVELREAAIF 994  
 DB 11 SEEKAAIVQSAALCCVRSKITQ-YEAVQSEGVMIQEWYKASGLACSOAEVHSSRAVY 69  
 QY 995 IQRIWRRRLFAKLLAAAEATARLQRSQKQAAASVIO--MOWRTYOLGRIOHREFLRQD 1052  
 DB 70 IQKAF-CRMVTRKL-----ETQKCAALRIQFFLQMAVY-----RRRFVQOKR 110  
 QY 1053 LIMFVQRMRSKSNLQORKEFOOLKRAAINIQWRRAKLSMRKCNADYLALRSSVLKQV 1112  
 DB 111 AAITLQHYFRT-W---QTRKQFLLYRKAADVLLQNHYRAFLSAKHQORVYLQIRSSVIL 166  
 QY 1113 AYRKATIQMRIDRNHYSLKKNVICLQORLRAIMKREBENYLRNLSILVQ---KRY 1169  
 DB 167 ARSFGFIQKR-----KFEIKNSTIKIQAMWRY 195  
 QY 1170 RMRQMIQDRNAYLTRKCIINVQ---RRWRATLQMRERKNYHLQTTTKRIQIKPRAK 1226  
 DB 196 RAKK-----YLCVKAACKIQAWYRCWA-----HKEYLAVLKAVKIIQGCFTYK 240  
 QY 1227 REMKKQABEFLQKKVILVQKRRALLQMKERQEVHLHREVTIKLQRRPHAQKSMPFM 1286  
 DB 241 LE-----RTRFLNVRASAILIQKWRRAILPAKIAHEHFLMIKR-----278  
 QY 1287 RAKYRGTOAAVSCLOMHWNRHLKRENSFLQRLQAAITLQRRYARLNMIKOLKSYAQ 1346  
 DB 279 -----HRAACLIQAHYRG-----YKGOVSLRQ-----301  
 QY 1347 LKQAAITQIOTRYAKKAMQKVLYQKOREAIKQVRRYRGNLEMRKQIEYVQKORAVI 1406  
 DB 302 -KSAALIIQYTRAREAGKHERIKYIBFKGTVILQALVRGWLVRKRFLE-----QRAKI 355  
 QY 1407 RLQKWRISIRDMRLCKAGYRRIRLSLSIQKWRATVQARQOREIFLSTIRKVLMOAFI 1466  
 DB 356 RL-----LHPTAAAYHLN-----369  
 QY 1467 RATLLMRQORREFEMKRRAAVVIQRRFRACAMLKARQDYOLIQSSVILVORFPRANSR 1526  
 DB 370 -----AVRIQRAYKLYLVANKQV-----NSVICIQWFRARLOE 406  
 QY 1527 KQARQEVQIURTIAVHLQKFRGKRLMLEQRCNCFLLRCSMPGFOARARGMARKRFOAL 1586  
 DB 407 KRFIQKYHSI-----KKIEHGEQCLSQNRNRAASVIAQVHRELLRKK-----449  
 QY 1587 MTPEMDLIRQKRAKV--IQRYWGRYLIRRRQKHQGLDIRKRIAQLRQAKAVNSVRC 1644  
 DB 450 -----QKFTSGIIKIQALWRGYSWRKKNQDCTIKAILRSLQVNVNREIRENKLIYK 500  
 QY 1645 KVQBAVRFLRGFRFIASDALAVLSQDLRLSRVPHLLMWCSEFMGTFCYGVIMQAQAIRSEVD 1704  
 DB 501 RTALALHYLLTYKHLISAILEALKHLEVTR-----LSPLCCENNAQSAISKI 548  
 QY 1705 KQIERSCSR-----IILNARYNSTVTNTPQEGGLVTIAQMLL-----1742  
 DB 549 FVLIRSCNRSIPCEMIVIRYAVQVLINVSKEYKTKTSVAVYDVENCIDILLELLQIYREKPGN 608  
 QY 1743 RWCQKDSIFNTLCTLIWVFAHCPRKRIIHDYMTNPEALYMWRETKKLVARKEKMQWA 1802

```

Db 609 KVADKGGSIPTKCLLAILL---KTTNRASDVRSKVVDRISYLYKLTAHKHKM--NT 663
QY 1803 RXPPPMTSGRYSQK-----INFTPCS-----LPSLEDPFGIIRYSPYTFISSVAFDT 1851
Db 664 ERI-----LYKQKNSISIPPIPTVTRIVSRKLPDVLRRDNMBEETNPLOAIQM 717
QY 1852 ILCKLQI 1858
Db 718 VMDTLGI 724

RESULT 11
Q84228
ID Q84228 PRELIMINARY; PRT; 1110 AA.
AC Q84228;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNBa0036M16.11 protein.
GN OSJNBa0036M16.11.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBa0036M16.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005103; BAC56022.1; -.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF00307; CH; 1.
DR Pfam; PF00612; IQ; 5.
DR SMART; SM00185; ARM; 1.
DR SMART; SM00033; CH; 1.
DR SMART; SM00015; IQ; 4.
DR PROSITE; PS50021; CH; 1.
DR PROSITE; PS50096; IQ; 2.
SQ SEQUENCE 1110 AA; 125271 MW; 830CB31B3A57958D CRC64;

Query Match
Best Local Similarity 19.3%; Pred. No. 2.8e-15;
Matches 268; Conservative 212; Mismatches 487; Indels 425; Gaps 52;

QY 419 AISPCKQVEDTTLPRSAAPANASARSSAHAWPHAQSKFKLAQTMSLMKPATPKV 478
Db 7 AVSSPPFD-LSNURTPNPRAPPKSSAKKEPLP-----SATPTTR 47
QY 479 RDTSIQSVKLYDSELYMQTCINPDPAFTTTIDPFLASTMYLDEQ-----AV 526
Db 48 RRGPPPP-----PRCAATATPLARRALDLQSSARRAESGRDGL 92
QY 527 DRHQADPKWNLVALSIPA-----DLDA-DLNNKIDV-----GKLFNE 563
Db 93 RAFASATSLSLLLRDPSCAGCAPSAAARVTRDAPAHGVQGRDAVDGERARGSPKR 152
QY 564 VRNKELVVAPTKEEQSNYLTLYEL-ETLRKA-----AVELPFEQMRLPSCSVAVY--- 614
Db 153 HRGEGDRGGPPRRKTKTTPMAASLRDSLRVCSLDVDTVMRGMSHMSREACEEVLVMQC 212
QY 615 ---VNKQALIRIDRNLDHVVWQRTILELLICFNPLRLGLEVVFG-----EKIQW 665
Db 213 ICKNIINGRLKMECHCPVSDLRDKATIRIFWCYNPKWLRIGLHIVLGSDWIQNESQK 272
QY 666 SNRDIVLSTFIILNRLF-----RNKCEEQRYSKAVTLTEEYAEATIKKHSLOKILFL 716
Db 273 KDEVAFLEKQLFVHMNTQPSALNKAPEGHRAC-----YAEAAASNLKELFL 326

```

```

QY 717 LPFLDOAKCR-----IVKHNPCLFVKKSPHKETKOILLRFSSELLANIGDITRE 766
Db 327 VAALDRAKTESGLPSBSGIDGLDGGSPLLFCRTEIKSSRQIVQESLGEVHGEGDILMH 386
QY 767 LRLRGVLVLRQRTFDLDFDYAFNNLAVDLRDGVRLLFRVVEVILLRDDLTRQLRV--PAIS 824
Db 387 LNIIMGYKLVYQQLALSEYDFTVGNLFEDLDQDGIILCRIIQ--LLTSDASILKVIAPSDT 444
QY 825 RLQRIPIVVKLALGALGEANFOL-----GGDIAQDIDVGHREKTLISLLWLIYKFRSPKFH 880
Db 445 YKKRLHNTCTMAIQYIKAGFPLSDADGLSISAEIDVNGDKELILALLWNFIYMLPVLV 504
QY 881 AAATVLOKWWRRHVLHVLRIRHKELMRHRAATVIAVFRGHQMRXVVKFKTERTQ 940
Db 505 NETSVAQ-----EISRLK 517
QY 941 AAAILQKFTRRYLAQKLYQSYHSIITIQRWRAQQLGRQHRQFVELREAAIFLQRIWR 1000
Db 518 APVSEQSISEMKSQTGLLYD-----WI 539
QY 1001 RLIFAKKLLAAETARKLQSQKQAAAASYIQMWRTYQLGRQRFELRQDLIMFVQRR 1060
Db 540 QVVCAYKGISVSSSQIDRRALNYFISYILININIPNPL-----578
QY 1061 MRSKWSMLRQKEFQQLKRAAINIQORWRAKLSMRKCNADYLALRSSVLKVQAYRKATIQ 1120
Db 579 ---KETLSDCRKELFSCHTDIADITTYQFNKIDLPFGWNLANDVLFVE--KSAIIL 633
QY 1121 MRIDRNHYYSLRKNVITCLOQLRLAIRMKREORENYL--RLRNASILVQKRYMRQMIQD 1178
Db 634 LAFSLSHLNVRR-----LEQLKNLIDSKLDHQSIVTE-----666
QY 1179 RNAYLTRKCIINVQRWRATIQMRERKNYHLQTTTKRIQIKFRAKRMKKQRAEFLO 1238
Db 667 -----VSPRRSRGTTDMK-----CHFPQTEE--TDGSRSTREW-----698
QY 1239 LKKVTLVOKRRALLQMKERQEVYHLREVTIKLQRRFHAQ-----KSMRFMRKYRG 1293
Db 699 ---AATVIQOTQARRLNAMSK---YCKLNATQPCNKGHDPASSPLKSIADSSCDSAT 751
QY 1294 Q-----AAVSCLOMEHNRHLKRRRNSFQLRQ--AAITLQRRYRARIINMIKOLK 1342
Db 752 KLVCEDDVDCCSSNCQVLYHDPVSTKVD---FLFCEKAWAARKIQFAYRRFAHRISRI 808
QY 1343 STAQLKQAATITQT-----RYRAKAMQKVLYYQKOREALIKVORRYGNLEMRKQI 1395
Db 809 S-----AAIKIQSHWRCSFVIRFKRQIQNTTIOAVARTSAISVQSFVRGWL-IRKQV 861
QY 1396 EYVQKQRAVIRLOKWWRSIRDMRLCKAGVRRIRLSLSIQ--RKWRATVQARRQREIF 1452
Db 862 ---KQILCSVYLIQRRWQVLFESKR-----SVIVIOAHVRGWIARQTVARNK--- 908
QY 1453 LSTIRKVRIMQAFIRATILMQORREFEMKRAAIVIORFRFARCAMLKARQDYLIQSS 1512
Db 909 ---KXITIIQSVKAYLLRKSKQBEITDDIMHMI---OKLRVNC-----FRLVSPG 953
QY 1513 VILVQKFRANKSMQARQOEVLQRTIAVHLQOKFRGKLMTEQRCNFOLLRCSMGFGQA 1572
Db 954 T-----ATEHSEKCCOTIVNAGAVEILLQINLLNRGVPDQEVLLKQVL----- 996
QY 1573 RARGEMARK--RFOALM-----TPEMMDLIQKRAAKVQRYWVRGYLIRRKQKHQGL-- 1623
Db 997 ---FTLRINARFNLQPVLIANTPOAVEIVFQELL-----RSTGEFFVA 1037
QY 1624 -DIRKELIAQLRQEAKAVNSVR-----CKVQEAHVFL-----RGRF-IASDALAVLSOL 1669
Db 1038 CHILKELCESSEGHKATARKHIRLGLCLVQLEKKVDLDKNGTGTAKD-----NNL 1092
QY 1670 DRLSR--TVPHL 1679
Db 1093 RRLGEAVTLHL 1104

```



RP	SEQUENCE FROM N.A.
RA	Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,
RA	Schuelker C.,
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RA	Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project;
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL021890; CAAL7149.1; --
DR	EMBL; AL461556; CAB79137.1; --
DR	PIR; T05464; T05464.
DR	InterPro; IPR008938; ARM.
DR	InterPro; IPR000225; Armadillo.
DR	Pfam; PF00612; IQ; 4.
DR	SMART; SM00015; IQ; 3.
DR	PROSITE; PS50176; ARM_REPEAT; 1.
DR	PROSITE; PS50096; IQ; 2.
KW	Hypothetical protein.
SQ	SEQUENCE 1088 AA; 123349 MW; D56F705E088E595E CRC64;
	Query Match 3.4%; Score 324.5; DB 10; Length 1088;
	Best Local Similarity 18.9%; Pred. No. 2.7e-12;
	Matches 219; Conservative 169; Mismatches 387; Indels 383; Gaps 42;
QY	388 SPVKNPKRSHSLHSFSDAPNESLYRNETHVAISPPKKQRVEDTTLPRSAAPANASARSS 447
Db	: :       :
Db	12 APPNPASSLLDSINSKPTERTSVV-NSNISKSPYPHFHTASKQTPEKSS--SNFRPSP 68
QY	448 SAHAHPAQS-----KKPKLAQTWSLMKMPATPRXVRDTSIQPSVKLYDSLVMQTC 499
Db	: :
Db	69 MVHSYASRSKVSTSSRLKAFELQQSOSRRKAALTKEK----- 106
QY	500 INPDFFAATTIDPFPLASTWYLDEQAVDRHQADFKKWLNALVSIPADLLDAD-LNNKIDVG 558
Db	_ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ : _ :
Db	107 -NURSLAKSLTV-----WLNFLEFPENCGCPFFNEGSGV 141
QY	559 KLFEVR-----NKELIVAPT-----KEEQSMVLTLYRIETLRKA 594
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	142 NLGKGKDGSEALGNKSXGVDTWWRSPKRLNLGWCEKERSIDSLTGSKYSTLR-- 199
QY	595 AVELFFSEQMRLPOS-----KVAYVNQAIRSRDNLIH 629
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	200 -----ESLRDVCSLDLKQRMQFHLSLGSCKEIFDVTMRVRSKNIDEGRIKKMQCPLV 252
QY	630 LDVVMORTILELLCFNPLMLRLGLEVVFG-----EKIQMSNDIVCLSTFILNLFRL- 683
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	253 TDFGMKEKALKMSYNQMVURLGLYIIFGDGSFLSDSEVNSDQMAFLKMWISKQFFSH 312
QY	684 -----NKCEEQRYSKAYTLTERYAETIKHSLOKILFLLPDLQAKOKRVK---- 730
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	313 DGLARSVYVKNWVEGLYRPG-----YYEALGNVILKRILLVLVIDRAKSQCCLSKYG 366
QY	731 -----HNPCLFVKXSPHKETDKILLRFSSELLANIGDIITELRRLVGVLOHRTFDDEF 784
Db	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	367 IDGIDGGSLMFSEKSIKGS-----HOLICGSEF 396
QY	785 DYAFNNLAVDLRDGVRLTRVVEVILLRDDLTRQLRVPAISRLORIENVKLALGAHGAFP 844
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	397 D-----LMRVLIKQKMVFSDNKKKNIANCRIALQYLKDAGV 433
QY	845 QLGGD-----IAAQDIVDGHREKTTLSLLQMLTYKFSRKPHAAATVLOKWWERHLHWVIQ 900
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	434 SLKDDCGWMITGEDVADGRELTITSLWNIFVHLQLP-----LLIN 474
QY	901 RRTPHKELMRHRRAATYIOAVFRGHOMRYKVLKFTKTQTAII-----OKFTTRY 952
Db	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db	475 GRLLTEIYK-----VOGYEQNNQI-----TWSTPLEMLINWTIDPGGGQGQSQV 519







*This Page Blank (uspto)*



```

Db 648 RHLLKSEBQERRHQLRREOQERREQLRKREBERLEQLKREHE-----EERRE 701
Qy 951 RYLAQKOLYQSYHSIT-IQW-WR-----AQQLG----- 978
Db 702 QELAEEOQARERIKSRIPKQWQLESEADARQSVLLLEAPQAGRAEAPQOEKERRRE 761
Qy 979 -----RQHRQRFVELREAAIFLQRIW-----RRLFPAKLLAAAEATLQRS 1020
Db 762 SELQWEEERAHROOQEEBQRDDFTWQ--WOAEKSEKRGRLSARPPLEQREQLRAE 819
Qy 1021 OKQAAAASYIQWRTYQLGRIQRH-----EFLQRDLIMFVRRMRKWSMLQ--RKE 1073
Db 820 ERQOEQRFLEPEEKEQGRQRREKELQFLEEEQL---QRRERAQQLQEEEDGLQE 876
Qy 1074 FQQLKRAAINIQORWRAKLSMRKNADYALRSSVLKQVAYRKATIQWRIDRNHYSLRK 1133
Db 877 DQERRRQQRDDQKRWQLEBERKRRH-----TLYAKPALQEQ-----LRK 918
Qy 1134 NVICLQQLRAIMKMRQRENVLRNASILVOKRYMRQOMIQDNAYLFRKCIINVQ 1193
Db 919 EQQLQEEEBEL--QREERKRRQEQ-----EYQYREEEQQLQEEEBELREER---EKR 968
Qy 1194 RWRATLQMRERKNYLHLOTTTKRI-----QIKFRAKREMKQRAEFLOLKKVTLVVQK 1248
Db 969 RQREERQYRKDKK---LQKEEQLLGEPEKRRQREKKYREEBELQEEELQEE 1024
Qy 1249 RRRALLQMRKEROEYLHREVITIKLORFPAKSMRFRKRVGTQAAVSCLOMHWNRHL 1308
Db 1025 RE-----KRRQEW--ERQYRKDELOQEEEBELREBERKRR-----LQERQY- 1067
Qy 1309 LRKERNFLQRAAITLQRYRRLNMVILKQSVYALQKAAITI-----QTRYRAKKA 1363
Db 1068 ---REEBELQEEEBELREBERKRRQREQLRQYRKREBELQEEEBELREBERKRRQERE 1124

RESULT 2
US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fredrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-644-94

Query Match 2.9%; Score 277; DB 2; Length 1898;
Best Local Similarity 21.7%; Pred. No. 5.6e-14;
Matches 188; Conservative 155; Mismatches 301; Indels 224; Gaps 44;

Qy 891 RRHLLVVIQRIHKLMMRRRAATVIAVFRGHQMKYVVKLFKTEQTAAILQKPTR 950
Db 648 RHLLKSEBQERRHQLRREOQERREQLRKREBERLEQLKREHE-----EERRE 701
Qy 951 RYLAQKOLYQSYHSIT-IQW-WR-----AQQLG----- 978
Db 702 QELAEEOQARERIKSRIPKQWQLESEADARQSVLLLEAPQAGRAEAPQOEKERRRE 761
Qy 979 -----RQHRQRFVELREAAIFLQRIW-----RRLFPAKLLAAAEATLQRS 1020
Db 762 SELQWEEERAHROOQEEBQRDDFTWQ--WOAEKSEKRGRLSARPPLEQREQLRAE 819
Qy 1021 OKQAAAASYIQWRTYQLGRIQRH-----EFLQRDLIMFVRRMRKWSMLQ--RKE 1073
Db 820 ERQOEQRFLEPEEKEQGRQRREKELQFLEEEQL---QRRERAQQLQEEEDGLQE 876
Qy 1074 FQQLKRAAINIQORWRAKLSMRKNADYALRSSVLKQVAYRKATIQWRIDRNHYSLRK 1133
Db 877 DQERRRQQRDDQKRWQLEBERKRRH-----TLYAKPALQEQ-----LRK 918
Qy 1134 NVICLQQLRAIMKMRQRENVLRNASILVOKRYMRQOMIQDNAYLFRKCIINVQ 1193
Db 919 EQQLQEEEBEL--QREERKRRQEQ-----EYQYREEEQQLQEEEBELREER---EKR 968
Qy 1194 RWRATLQMRERKNYLHLOTTTKRI-----QIKFRAKREMKQRAEFLOLKKVTLVVQK 1248
Db 969 RQREERQYRKDKK---LQKEEQLLGEPEKRRQREKKYREEBELQEEELQEE 1024
Qy 1249 RRRALLQMRKEROEYLHREVITIKLORFPAKSMRFRKRVGTQAAVSCLOMHWNRHL 1308
Db 1025 RE-----KRRQEW--ERQYRKDELOQEEEBELREBERKRR-----LQERQY- 1067
Qy 1309 LRKERNFLQRAAITLQRYRRLNMVILKQSVYALQKAAITI-----QTRYRAKKA 1363
Db 1068 ---REEBELQEEEBELREBERKRRQREQLRQYRKREBELQEEEBELREBERKRRQERE 1124

```



QY 1208 ---NYLHOTTXKRIQIKPRAKEMKQRAEFLOLKKVTLVQKRRALLQMKERO--B 1262  
Db 1572 ELTN--RLQBSQBEIQIMKEKEMKVQ--BALQIERDQL--KENTKEIVAKMKESEKE 1626  
QY 1263 YLHREVTIK--LQRRPHAQK--SMRFRAKYRGTOAAVSCLOM 1302  
Db 1627 YQFLKTAVNTEQKCEIEHLKEQFQTKLNLENIETENIRLTQILHENLEEMRSV-- 1693  
QY 1303 HWRNHLKREPERNSFLOLQRAAITL--QRRYRANIMIKOLKSYAOLK-- 1348  
Db 1684 ---TKERD---DLRSVEETLKVVERDQLENLTLAQLHENLEEMKSVMKERDNL--RRVE---ETL 1732  
QY 1349 QAAI---TIQTYRAKAKOKVYLYQKREAI---IKVORRYR--GNLEMEKOLEVY 1398  
Db 1733 QETIDKLRGIVSEKTEINNMOKDL---EHSNDALKAQDLKIQBELRIAHHLKEQOETI 1789  
QY 1399 QKQRAV---IRLOKWRGIRDMRLCKAGYRIRLSLSIQRKWR 1440  
Db 1790 DKLGIYSEKTKLSNMOKDLENSNAKLOE---KIQELKANEHQILITLKDVNETQKVS 1846  
QY 1441 ATVOARRQREIFLSTIRKVRIMQAFIRATILLMQORREFEMKRRAAVVIQRRFRARCML 1500  
Db 1847 EMEQLKKQIKDQSLTSLKLEITENLNAQELHENLEEMKSVMKERDNL--RRVE---ETL 1900  
QY 1501 KARODYOLIQSSVILVQKFRANRSMKQARQEFVQLRTIAVHLOQKFRGKELMIE--QRN 1558  
Db 1901 KLERD-QIKESLOETKARDLBIQELKTARMLSEKHEKTVDKLREKISEKTIQISDIQKD 1959  
QY 1559 CFQLLRCSMPQFQARARFARKFQALMTPEMDLIRQK--RAAKVIQRYWRGYLIR-- 1614  
Db 1960 ---LDKSKDLQKIQE--LOKKELOLLRVKEDVNMHSHKKNEMQLKKQPEPNVLCKE 2014  
QY 1615 ---RQKHQGLDIRKRIAQIQEAKAVNSVRCKVQEAHVRFLRGRFIASDALAVLSQ 1668  
Db 2015 MDNFQTKLHESLEEIR-IVAKERDELK-----RIKESLKMERDQFIAT--LREMLA 2064  
QY 1669 LDR-----LSRTVPHLMWCSSEFMFCYGINAQAIRSEVDKQLIRCSRIILN 1717  
Db 2065 RDRQNHQVPEKRLISDCQOHLH-----ESLREKCSRIKEL 2100  
QY 1718 LARYNSTVTNTFOEGLVTIAQMLRLWCDDSEIFNTLCTLIWVFAHCPKRRKIHDYMT 1777  
Db 2101 LKRYSE-----MDHYECLNRLSLD--LEKEIEFHRIMKLLKYVLVSYVTKIEBQHECIN 2153  
QY 1778 NPEAIYM--VRETKKLVARKEMKQNAKPPMTSGRYKSOKIN-----FTP 1822  
Db 2154 KFEMDFIDEVEKQKELLIKIQHLOQDCDVP---SRELRLKLNQNDLHIEITLKDFSE 2209  
QY 1823 CSLPSLEPDF 1832  
Db 2210 SEFPSIKTEP 2219

## RESULT 4

US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

## Query Match 2.8%; Score 268; DB 1; Length 2482;

Best Local Similarity 18.0%; Pred. No. 5e-13;  
Matches 299; Conservative 294; Mismatches 638; Indels 428; Gaps 69;

QY 105 HPPQAPLVEKNVYK-----TPQEPVVISPOPSRLKENLSMTP--GNLL 148  
Db 745 HPVSLAPLDESNSYEHLLTSLDKEYMHFAELQKFLSLQSEHKILHDOHCQMSKMSLQ 804  
QY 149 DVIDNLRTPLITETRGGOATIFPDNLAAWPTPLK-----GNVKSANDMRPRIT 200  
Db 805 TYVDSL-----KAENLVLTNLFNFGDVLKEMQGLGRLGVLPSLSSSCVPS 854  
QY 201 PDDLED---QPATNKTFDVKHSETI-----NISLDTDCSRIDGQHPPLANKTTIV 249  
Db 855 LSSLDGSSFYRALLQETGDMSSLLNLEGAVSANQCSVDVFCSSL----- 899  
QY 250 HATITRALACHHEGSPPTPTKSAIHLKRDIKLVGSLPKRYSESMDLSLLSPQTK 309  
Db 900 -----QENLTKETPSAPA-----KGVEELESCEVYRQSLK-----BEK 937  
QY 310 YAIQSGMENLEMKIRSTEQ-----NRYQBOQIQIKAKDLNLSSSSEASLAG 357  
Db 938 MESQGIWKN---KEIQELEQLLSERQELCLRKQYLSENEQWQOKLTSVILEMESKLA 994  
QY 358 QOEFNFHSEILAQSSRFNLH--EYGRKSVKSGPVKNPKRSHLSFSDAPSNESLYRNE 416  
Db 995 EKK-----QTEQLSLEVARLQLOGLDLS--RSLIGIDTDAIQG-----RNE 1037  
QY 417 TVAISPPKQKQVEDTTIPRSAPANASARSSAHAWPHAQSKFKLAQTMIMKPP---A 473  
Db 1038 SCDIS---KEHTSETT---ERTPKHDVHQICDQAQDLNLDIEKITETGAL--KPTGEC 1089  
QY 474 TPKVRDTSIQPSVKLYDSLYMOTCINPDFFAATTTIDP--FLASTMYLDEQAVDRHQ 531  
Db 1090 SGEQSPDTNYEPGGE--DKTQGSSECTISELFSGPNALVPMDFLG-----NQEDIHNLQ 1142  
QY 532 DFKKWLN---ALVSIAPADLADLANKIDVGKLFNEVRNKLIVAPTEEQQNNVYTKYRL 588  
Db 1143 RVKETSNNENRLLHVIEDRRD-----KVESILNEM-----KELDSKLHLQEVOL 1186  
QY 589 ETLKAAVEL-----FSEQMR-LPCSKVAVYVNVKQALR--IRSRNHLHLDVVM 634  
Db 1187 MTKIEACIELEKIVGELKKENSJLSEKLEFYSCDQELLQRVETSEGNLSLEMAADKSS 1246  
QY 635 QRTILELLLCNPLWRLGLEV-----VFGEKIQMSQNRDVIGLSTFILNRL 681  
Db 1247 REDIGDNVAKYNDKWKERFLDVENELSRIRSEKASIEHEALYLEADLEVQTEKLCLEKD 1306  
QY 682 FRNK-----CBEQR-----YSKAYTITTEYAEATIKKHSLOKILFLFPF 719

Db 1307 NENKQVIVCLBEELS SVT SERNQLRGELDTMSKTTTALDQSEKMK 1354  
QY 720 LQQAQKRVKXNP-----CLFVKKSPHKYKIDILLRPFSSLLANIGDITELRRLGVVLO 775  
Db 1355 -----KTQELSHQSECHICQIAEAEVKEKTELLQTLSSDVSLDKDTH-----LQ 1402  
QY 776 HRQTFDEFDYAPNNLAVLDRCGVRLTRVVEVILLRDDLTRQLRVPALSRLORIENVKLA 835  
Db 1403 EKLOSEKDSQALSLLTKCELENGIAQINKEKELIVKESLSQARL-SESDYELK-NVSKA 1460  
QY 836 LGAL-----GEANPOLGGDIAAQDIPVGHREKTLSSLWLIYKFRPKFPHAAATVLOKWMR 891  
Db 1461 LEAALVEKGEFALRLS---STQBEVHQLR-RGIEKLRVRIEADKKQLHIAEKLERERE 1516  
QY 892 RHWLHVVIQRRIRHKELMRHRAATVIOAVFRGHQMKYVKLTKTERTQAAITLOKFTTR 951  
Db 1517 NUSLKDKNLEKLELQSENGEQLVILDA-----ENSKAEVETLKTQTEEMARSLKVPFL- 1571  
QY 952 YLAQKQLYQSYHSIITIQRWRAQQLGRHQRFVELREAAIFLQRIWRRRLFAKLLAA 1011  
Db 1572 -----DLVTILRS--EKENLTQIOEKQGOQLSELDKLLSSF-----KSLLEE 1610  
QY 1012 AETARLQSQKQAAASYTQMOWRTYOLGRI-----ORHEFL 1048  
Db 1611 KEQAEIQIKESKSTAVEMLQNLK--ELNEAVALCGDQEIIMKATBQSLDPPIETEEHQL- 1667  
QY 1049 RORDLIMFQRM-----RSKWSMLEQKEFQO---LKRAAINIQRWRAKLSMRKNAD 1100  
Db 1668 ---RNSIEKURALEADEKKQLCVLOQKSEHADLLKGRVENLR-----ELBIARTNOE 1721  
QY 1101 YLALRSSVLK-----VQAYRKATIOMRIDRNHYSLRKNVICLOQLRAIMK 1147  
Db 1722 HAALAEKSGEVETLAKIEGTQSLRGLELDVITIRSEKENLTNLOKQERISELEI 1781  
QY 1148 MREQRNYLRLNASILVQKRYMRQ-----QMIQDRNAYLTRKCI--NVQRRWRATL 1200  
Db 1782 INSSPENILOEKE-----QEKVQMEKSSSTAMEMLQTLKELNERVAALHNDQEAACKAE 1836  
QY 1201 Q-----MRRERKNYLIHQTTK-RTOIKFRAKREMKQBAEFLOLK 1240  
Db 1837 QNLSQVECLELEKALLOGLDEAKNNYIVLOSSVNGLIQEVEDGKQLEKDEIRSLK 1896  
QY 1241 KYTLVVQKERRALLQMKERQ-----EYLHLREVTIKLQRRFHAQKSMRFMAKYRGTQAA 1296  
Db 1897 NQIQDEQLVSKLSQVEGHEHQLKQENLELRLNLTVELEK-----IQVLQSKNASLQDT 1950  
QY 1297 VSLQMHWRN-----HLLRKERNSTFL-----QLRQAAITLQRRYAR 1334  
Db 1951 LEVLQSSYKRNLELELTAKMDKMSFVEKVNKMTAKETELQREHMAQKTAELQEBLSGE 2010  
QY 1335 LN-----MIKQKLSYA-OLQQAATITQYRAKKAQK-QVVLVYQKOREALIKVQRR 1384  
Db 2011 KURLAGELQLLEEEKKSDKQLEKLTLENSLCKSLDCMHQDQVEKEGVREIEAYQLR 2070  
QY 1385 YRGNLEMRKQIEVYQKQAVIRLOKQWRSIRDMRLCKAGYR-----RIRLSLSIQKWR 1440  
Db 2071 LH-EAEKHQALLDNTNKQYVEIQYREKLSKECLSSQKLEIDLLKSSKEELNNSLK 2129  
QY 1441 ATVQARRQREIFLSTIRKVRMLMOAFIRATLMLRQORREFEMKRAAVVIOFRFRACML 1500  
Db 2130 ATTO-----LLEELKTKMDNLKYNVQNLKXENBERAQKMK-----LLIK-----SKQL 2173  
QY 1501 KARQDYQLTOSSVILVORFRANRSMKQARQFVQLRTI 1539  
Db 2174 E--EKEILOKEL-----SOLQAQAEQKQGTGTV 2199

## RESULT 5

US-09-976-594-736

; Sequence 736, Application US/09976594

; Patent No. 6673549

; GENERAL INFORMATION:

; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 736  
; LENGTH: 1530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Inocyte ID No. 6673549 3151579CD1  
US-09-976-594-736

Query Match 2.8%; Score 264; DB 4; Length 1530;

Best Local Similarity 18.5%; Pred. No. 5.1e-13;  
Matches 294; Conservative 277; Mismatches 517; Indels 498; Gaps 69;

QY 263 EEGPSP-----RTPTKSAIHLKR---DIKLVGSPLRKYSEGMKDL 301  
Db 68 QNGPTFPDPSPSLDPTTSPVGPDPASPGVAGPHDNLKRSQGTSAEGSVRKEALQSLR-L 126  
QY 302 SLLSPOTKYAIOGS-MPNLNMKIRIBONRYVYRQOIQIKAKDLNSSSSEA----- 353  
Db 127 SLFMQETQLCSTDSPLEKEQVR-LOAKMLEQLKQVRKVRKQCRSSQPAKTRLFS 185  
QY 354 -----SLAQOQEFLEHSEI-----LAOSSRENLHEVGKSVKSPVKN 392  
Db 186 TLDPELMPENLPRASTLANTKEYSFLRTSPRGKVGSLGLPAHPREKTKSRSKIRS 245  
QY 393 PHKRSHELFSFAPSN-----ESLYRNET-----VAISPPKQORVEDTTLR-R 435  
Db 246 LADYRTEDSNAGSGVNPAPDSTKGLKQNRSSAASVSEISLSDTDRLENTSAGD 305  
QY 436 SAAPANASARSSAHAWPHAQSKFKLAQMSLMKPAKPRKVRDTSIQPSVKLYDSLEY 495  
Db 306 SVSEVDGNDSDSSYSSASTRGTYGILSKTVGTQDTF-----Y 343  
QY 496 MQTCINPDPFAATTIDPPLASTWYLDQAVDBHQADPKKWLNALVSIADLDDADLNKI 555  
Db 344 M---VNGQIPADT-----LGQFFSIKDVLOQAAAAEHQ 373  
QY 556 DVGKLFN-EVRNK-----ELVVAPTKEEQSMNYLTKYRLETIRKAAVELPFSEOM 604  
Db 374 DQGEVNGEVRSEKSDSICSSVSLESSAAETQEM-----LQVLK-----EKM 415  
QY 605 RLPCSKVAVYVN-KQALRIRSDRNHLHDVVMORTILELLCFNPLMLRLGLEVVFEKIQ 663  
Db 416 RLEGQLEALSLEASQALK-----EKA 437  
QY 664 MOSNRDVLGSLTFILNLFKRNKCEQRYSKATLTTEEYAEITKHK--SIQKILFLFLPFD 721  
Db 438 LQA--QLAALSTKQAOV---ECSSHSSQORQOSLSSE-VDTLKQSCWDLERAMTDLQNNL 491  
QY 722 QAKQKRIKVNHPCLFVKKSPHKETKDILLRFSSELLANIGDITRELRRLGYVLQHRQTEL 781  
Db 492 EAKNASLASSNNDLQVABEQYQ-----RLMAKVEDMQRM----- 526  
QY 782 DEFDYAFNNLAVDLRDGVRL--TRVVEVILLBDDITRQLRV--PAISRLORI---FNVKL 834  
Db 527 -----LSKDNVTVDLRQOMTALQSQQLQQLERTTLTSLKLSQAQABISSLQSVRQWYQQL 582  
QY 835 ALGALGEANFOLGGDIA-----AQDIVDGHREKTLISLLWOLI-YKFRSPK----- 878  
Db 583 ALAQ--EARVLOGEWAHIQVGMQTQAGLLEHLEKLENVLSQQLTETQHRSMKEGRITAA 640  
QY 879 -----FHAAATVLOKWWRRHVLHVVIQRRIRHKELMRHRAATVIOAVFRGHQMK 929

Db 641 QLQIEADMLDQAAFWOIQEAKTWBDEDLQRLLEPE-GERERLQRMADSAASLEQOLE 699  
 Qy 930 YVKLFKTERTOAAIILOK-----FTRYIAQKQYQSYHSIIITQWRAQOLGRHOR 984  
 Db 700 QVKLTLLQDQALQOEHLDMKQTLTQEAQSQESQSLDALQTHYDELOARLBLOQ 759  
 Qy 985 FVELREAAIFLQIRWRRLFAKLLAAAEATAR-----LQSQKQOQ 1024  
 Db 760 BAASREDTICL--LONEKIIIEAALQAQKSGKBELEDGARRLEBGTETSETLEKLEEL 817  
 Qy 1025 AAAS-----YIQMWRTY--QLGRQIRHEFLQRDLIMFVQRMKSKWMLQKBFQOLK 1078  
 Db 818 AIKSGQVEHLQOEFAALKKQWOKI-KEQFLQOKVMVEAYRRDATSKQOLISE----- 868  
 Qy 1079 RAANIQORWEAKLSMRKCNADYIALRSSVLKQVAYRKATIOMPIDRNH--YYSLRKNVI 1136  
 Db 869 -----LKAIRKELDSSELKELRQELQMVHVG-EKRTAEABELSLHREVAQVROHMA 916  
 Qy 1137 CLOQLRAIMKWRQORENYLR-----LKNASILVQKRYMRQMIQDRNAVYLR 1185  
 Db 917 DLEGHLSAQKEDERMETHLQSLQFDKEQMVAVTEANEALKQIEELQO-----EA 967  
 Qy 1186 RKCIINVQWRW-----ATLQMRERKNYLH-LQTTKRIQIKPRAK-----EMKQ 1232  
 Db 968 RKAITQOKMRRLGSLDLSAQKEMKTKHAYENAVGILSRLOEALAAKEADAELGOL 1027  
 Qy 1233 RA-----EFLQKKVTLVQGRREALQMKERQOYLHL-----REVTIK 1272  
 Db 1028 RAQGSDDSSALHERIOALEABLOAVSHSKTLE--KEIQEVIALTSQELSESEKYLE 1085  
 Qy 1273 LQRRFPAQSMRFRKRYRGVQAQVSCIQMWRNHLKRRNSFLQRLQAAITLQRRYR 1332  
 Db 1086 LEDEL--QESRGF-----RKKIKLEBSN-----KKLALBLEHE-K 1118  
 Qy 1333 ARLNMIKQLSYAOLQAQAAITQIRYAKKA-----MOKOVLYQKQREAIIVQRRYR 1386  
 Db 1119 GKLTGLQ--SNAALREHNSILETALAKREADLVQLMLQVAVLQKKEE----- 1166  
 Qy 1387 GNLEWRQIYVQ--KQOAVIFLOKWSIRDMLRCAGYRIRLSLSIQKWEAT 1442  
 Db 1167 -DROMKHLVQALQASLEKEKVSNSKEQVAAK-----VEAGHNRHFKAAASLE- 1215  
 Qy 1443 VQARRQREIFLSTIRK-----VRLMQAFIRATILMQORRBFEMKRRAAVVIQRRF 1494  
 Db 1216 -----LSVYKELQAKELHVQKLOA-EADDLQIREGKHQSEIAQFOAELAEA--R 1262  
 Qy 1495 ARCAMLKARQYQIQLQSSV-----ILVQKFRANRSMK-----QARQEFVQIR 1537  
 Db 1263 AOLQLQQLQDEQLSKQPVGNQEMENLKWEVDQKERTQSLKQQLDLTEQOQGRKELEGLQ 1322  
 Qy 1538 TIAVHLQOKFR--GKRLMIEQRCFQILRCMPGFOARARGFMARKRFQALMTPEMDLI 1595  
 Db 1323 QLLQNVKSELEMAQEDLSNTQDKFPM-----QAKV-----SELKNNKNTLL 1364  
 Qy 1596 RQKRAAKVIQRYWRYGILRRQKHQ 1621  
 Db 1365 QONQQLKD-----LRRQDEKG 1382

## RESULT 6

US-08-353-700-1

; Sequence 1, Application US/08353700

; Patent No. 5599919

; GENERAL INFORMATION:

; APPLICANT: YEN, TIMOTHY J.

; APPLICANT: RATTNER, JEROME B.

; TITLE OF INVENTION: NUCLEIC ACID ENCODING A

; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

; TITLE OF INVENTION: AND METHODS OF USE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN

; STREET: 1601 MARKET STREET, SUITE 720

CITY: PHILADELPHIA  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-2307  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-1

Query Match 2.7%; Score 259; DB 1; Length 3248;

Best Local Similarity 18.3%; Pred. No. 4.5e-12;

Matches 268; Conservative 271; Mismatches 572; Indels 356; Gaps 62;

Qy 262 EEEGSPRRPTKSAIHDIKLVGSPURKYSGMKDLSLLSPQTKYAIQSGMPNLNE 321

Db 1628 QEENLTRKETPSADA-----KGVELESCLCEVYRQSLK-----EKKMESQGIKMN--- 1674

Qy 322 MKIRSIQ-----NRYOEOQIQIKADLNSSSSSSSEASLAGQQEFLNHSIL 369

Db 1675 KEIQEQLLSERQELDCIRKQYLSNEQOQKLTSTVLEWESKLAEEK----- 1725

Qy 370 AQSRNMLH-EVGKSVKGVKPNPKRRSHLSFSADPSNESLYRNETVAISPPKKQV 428

Db 1726 -QTQLSLELVARLQQLDLS--RSLIGITDEAIOG---RNESCDIS---KEHT 1774

Qy 429 EDTTLPRSAAPANASARSSSAHAPHAQSKFKKLAQTWSLMKKP---ATPRKVRDTSTQP 485

Db 1775 SETT---ERTPKDHVHQICDKDAQDNLMDTEKITETGAV--KPTGECGSGQSPDTNYP 1829

Qy 486 SVKLYDSLYMQTCINPDPPRAATTIDP--PLASTMYLDEQAVDRHQADFKNLNL---AL 540

Db 1830 PGE--DKTQGSSECSLSLSPGPNALVPMDFIG-----NQEDIHNLQRLVETSNENURL 1882

Qy 541 VSIADLADLNNKIDVGLKLENEVRNKLTVVAPTKEQSMNYLTKYRLETLKKADEL-- 598

Db 1883 LHVIEDRR-----KVESLLNEM-----KELDSKLHQEVQLMTKIEACIELEK 1926

Qy 599 -----PFSEQMR-LPCKSVAVYVNNKQALR-IRSDRNHLHDVVMQRTILELLLCFN 646

Db 1927 IVGELKKNSDLSEKLEYFSFCHQELLQRLVETSGELNDEHADKSRREDIGNVAKVN 1986

Qy 647 PLWRLIGLEV-----VFGEKIQMSNRDIVGLSTFTILNRLFRNK-----CEE 688

Db 1987 DSWKRFIDVENELSRIRSEKASIEHEALYLEADLEVQVOTEKICLEKNENKQKVIUCLE 2046

Qy 689 QR-----YSKAYTLTEEYATTKGHSLOKILFLPLFDQAKQKRIYKH 731

Db 2047 EELSVTTSERNQLRGELDTMSKTTALDQLSEKMK-----KTOELES 2090

Qy 732 NP-----CLFVKKSPHKEKDIILLRPFSSSELLANIGDITRELRLGYVQHRQFLDFDYA 787



```
Db 2091 QSECLHCIOVAEEVEKTELLQTLSSDVSELKDKTH-----LQEKLOSLKXDSQA 2142
QY 788 FNNLAVDLRGVRLTRVVEVILLRDDLTQLRVPALSRQRIENFKALGAL-----GEAN 843
Db 2143 LSLTKCELENQIAQLNKEKELLKESLQARL-SESDYEKL-NVSKALEAALVEKGEFA 2200
QY 844 FOLGGDIAAQDIDVGHREKTLISLLWOLLYKFRSPKPHAAATVLQKWRRHMLHVQRI 903
Db 2201 LRLS-----STQEEVHQL-RGIEKLRVRIEADKKQHLHAEKLERENDSLKDKVENLE 2256
QY 904 RHKELMRHRAATVIOAVFRGHQRMKYVLKFKTERTOAAIILQKFTYRYLAQKQYQSYH 963
Db 2257 RELQSEENQELVILDA-----ENSKAEVETLQIEEMARSLKIFEL-----2299
QY 964 SIITIQWRRAQOLGHRQRQFVELREAAIFLQRIWRRLFAKLLAAATAFLORSQK 1023
Db 2300 DLVTLRS--EKENLTQIOEKQOQLSELDKLLSPF-----KSLLEKEQAEIQIEES 2350
QY 1024 QAAASYIQMWRTYQLGRI-----QRHFRFLORDLIMFVORR 1060
Db 2351 KTAVELMQLQK--ELNEAVALCGDQEIKNKATEQSLDPIIEHQI-----RNSIEKLAR 2405
QY 1061 M-----RSKSMLEQRKEFOQ-----LKRAAINIQORWRKLSMRKCNADYLALESVVK-- 1110
Db 2406 LEADEKKQLCVLQQLKESBEHADLLKGRVENLER---ELEIARTQEHAALEAENSKE 2461
QY 1111 -----VQAYKATIQMRIDRNHYSLRKNVICLOORLRAIMKMRQENVILRL 1159
Db 2462 VETLKAKIGMTOSLRLGLDVTIRSEKENLTNELQEQERISELEIINSFENILOEK 2521
QY 1160 NASILVQKRYRMO-----QMIQDNNAVLTAKCII-NVORWRATLO-----1201
Db 2522 E-----QEKVQKEKSTANEMLTQOLKELNERNVAALHNDQEKAKQNLSSQVECLEL 2576
QY 1202 -----MRRERKNYLHQTITTK-RIQTKFRAKREMKKQRAEFLQLKKVTLVVKRRRA 1252
Db 2577 ERAQLLQGLDEAKNNYVQLSSVKGLIQEVEDGKQKLEKDEIRSLKNQIQDQEQOLVSK 2636
QY 1253 LLQMRKERO-----EYHLREVITKLORRHHQAKSMFRKRYGTQAAVSCLOMHWEN-- 1306
Db 2637 LQVQEGEHLQWKEQNLRLNLATVELEBQ-----IQVLOSKNASLODTLEVLQSSYKNLE 2690
QY 1307 --HLKFKRERNFL-----OLRQAAITLQRRYARLN-----MI 1338
Db 2691 NELELTQMDKMSFVEKVNKMTAKETELQREHMEWAQTAELQELSGEKNRLAGELOLLL 2750
QY 1339 KOLKSYA-OLKQAAITQTRYAKKAMQK-QVVLVYQKQRAIIVQVRRYRGNLEMRKQIE 1396
Db 2751 ERIKSKDQLKELTLENSELKSLDCKHKKDQVEKEGKVREEIAEYQLELH-EAEKKHQAL 2809
QY 1397 VYQKQRAVIRLQKWRSIRDMLCKAGYR-----RIELSLSLQKRWATVQARRQREIF 1452
Db 2810 LLDITNKQVEIQTQYREKTLSEKESLSSQKLEIDLLKSKKEELNLSLKATTO-----I 2862
QY 1453 LSTIRVRLMOAFIRATLMMQRRORREFFEMKRRAAVQIRFRARCAMLKARQDYOLIQSS 1512
Db 2863 LBELKTKMDNLKYVNLQKENERAQKKK-----LLIK-----SKQLE--EKEILQKE 2911
QY 1513 VILVQKFRANRSMKQARQEFVQLRTI 1539
Db 2912 L-----SQLQAAQEKQKTGTV 2927
```

```
RESULT 7
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match 2.7%; Score 259; DB 5; Length 3248;
Best Local Similarity 18.3%; Pred. No. 4.5e-12;
Matches 268; Conservative 271; Mismatches 572; Indels 356; Gaps 62;

QY 262 BEEGSPPTPTKSAIHLKRDIKLVGSPILRYSESKMDLSLSPQTKYAIQSGMPNLNE 321
Db 1628 QENLTKRTPSAPA-----KGVEELESCEVYRQSLEKL-----BEKMSQGIKMN-- 1674
QY 322 MKIRSEIQ-----NRYQEQQIQRKADLNSSSSSEASLAGQOEFNFHSEIL 369
Db 1675 KEIQLEQLLSERQBLDCLRKQYLSENEQWQOKLTSVTLEWESKLAEEK-----1725
QY 370 AQSSRFNLH-EVGRKSVKGVKPNKPHKRSHELFSDFAPSNESLYRNETVAISPPKKQV 428
Db 1726 -QTEQLSLEVARLQQLGLDISS---RSLIGDITDAIQ-----RNECDIS--KEHT 1774
QY 429 EHTLPRSAAPANASARSSSAHAWPHAQSKTKPKLAQTMSLMKP---ATPRKYVDTSIQ 485
Db 1775 SETT---ERTPKDHVQICDKDAQDNLNDIEKITETGAV--KPTGECGEGQSPDNYEP 1829
QY 486 SVKLVDSELYMOTCINPDFAATTIDP--FLASTWYLDQAVDRHQADFKNWLN---AL 540
Db 1830 PQE--DKTQGSSECSISELSFGFNALVPMDFLG-----NQEDIHNLQLRKETSNEHL 1882
QY 541 VSI PADLDADLNNKIDVGLFNEVRNKLWVAPTKEEQSMNYLTQVRLTETLRKAAVEL-- 598
Db 1883 LHVIEDRR-----KVESLLNEM-----KELDSKLHLOEVQVMTKEACIELEK 1926
QY 599 -----FFSEQMR-LPCKSKVAVYVKNQALR-IRSDRNHLVDVVMORTLELLCFN 646
Db 1927 IVGELKKNESDLSEKLEYSCDHQELLQVETSEGLNSDLEMHADKSSREDIGDNVAKVN 1986
QY 647 PLWRLGLGV-----VGEKIOMOSNRDVGISTFILNRLFRNK-----CBE 688
Db 1987 DSWKRFLLDVNELSRIRSEKASIEHEALYLEADLEVVQTEKLCLEKDNENKQKIVLCLE 2046
QY 689 QR-----YSKAYTLTEBYAETIKKHSIQKILFLLPFLDQAKQKEIVKH 731
```

Db	2047	EELS	VVT	SERN	QNL	R	G	E	L	T	M	S	K	T	A	L	D	O	S	E	K	M	K	E	-----	K	T	O	E	L	E	S	H																									
Qy	732	NP	-----	CL	F	V	K	S	P	H	K	E	T	O	I	L	F	S	E	L	L	A	N	I	G	D	I	T	R	E	L	R	L	G	V	I	L	H	R	Q	T	E	L	D	E	F	Y	A	787									
Db	2091	Q	S	E	C	H	C	I	O	V	A	E	B	V	E	K	T	E	L	L	O	T	L	S	S	D	S	E	L	L	K	D	K	T	H	-----	L	O	E	K	L	S	I	K	D	S	Q	A	2142									
Qy	788	F	N	N	A	V	D	L	R	D	G	V	L	T	R	V	E	V	I	L	L	R	D	D	I	T	R	O	L	R	P	V	A	I	S	R	L	O	R	I	F	N	V	K	L	A	G	A	L	-----	G	E	A	N	843			
Db	2143	L	S	L	T	K	E	C	E	L	N	Q	I	A	O	K	E	K	E	L	L	V	E	S	E	S	L	O	A	R	L	-----	S	E	S	D	Y	E	K	L	N	V	S	K	A	E	A	L	V	E	K	E	F	A	2200			
Qy	844	F	O	L	G	D	I	A	A	O	D	I	V	D	G	H	R	E	K	T	L	S	L	I	L	Q	I	L	K	F	R	S	P	K	H	A	A	T	V	L	O	K	W	R	R	H	L	V	I	O	R	I	903					
Db	2201	L	R	L	S	-----	S	T	O	E	V	H	O	L	R	-----	R	G	E	K	L	R	V	R	I	E	A	D	E	K	K	O	L	H	I	A	E	K	L	R	E	N	D	S	L	K	D	K	V	E	N	L	E	2256				
Qy	904	R	H	K	E	L	M	R	R	H	R	A	A	T	I	Q	A	V	F	R	G	H	O	M	K	V	Y	K	L	F	K	T	E	R	T	O	A	A	I	L	O	K	T	P	R	R	I	Y	L	A	O	K	Y	O	S	Y	H	963
Db	2257	R	E	L	Q	M	S	E	N	Q	E	L	V	I	D	A	-----	E	N	S	K	A	E	V	E	L	T	K	T	O	I	E	E	M	A	R	S	L	K	I	F	E	L	-----	2299													
Qy	964	S	I	I	T	O	R	W	R	A	Q	L	G	R	O	H	R	O	F	V	E	L	R	A	A	I	F	L	O	R	I	W	R	R	E	L	F	A	K	L	A	A	A	E	T	A	R	L	O	R	S	O	K	1023				
Db	2300	D	L	V	T	R	S	-----	E	K	E	N	I	T	O	L	O	E	K	O	G	O	S	E	L	D	K	L	L	S	F	-----	K	E	L	L	E	K	E	O	A	E	I	O	K	E	S	2350										
Qy	1024	Q	A	A	A	S	I	Q	M	O	W	R	T	Y	Q	L	G	R	I	-----	Q	R	H	E	L	R	O	R	D	I	M	F	O	R	R	-----	1060																					
Db	2351	K	T	A	V	E	L	Q	N	Q	L	K	-----	E	L	N	E	A	V	A	L	C	D	O	E	I	M	K	A	T	E	S	L	D	P	P	I	E	E	H	Q	L	-----	R	N	S	I	E	K	L	R	A	2405					
Qy	1061	M	-----	R	K	S	M	L	B	O	R	K	E	P	O	-----	L	K	R	A	I	N	I	O	O	R	W	A	K	L	S	M	R	K	N	A	D	Y	L	A	L	S	S	V	L	K	-----	1110										
Db	2406	L	E	A	D	E	K	K	O	V	L	O	I	K	E	S	H	A	D	I	L	K	G	R	V	E	N	L	R	-----	E	L	E	I	A	R	T	O	E	H	A	L	E	A	N	S	K	G	E	2461								
Qy	1111	-----	V	O	A	R	K	A	T	I	O																																															

APPLICANT: TOUTATLY, GARY  
 APPLICANT: LIDGARD, GRAHAM P  
 TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 STREET: 125 HIGH STREET  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/466,390  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PITCHER ESQ, EDMUND R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2101 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-466-390-4

Query Match 2.6%; Score 250; DB 1; Length 2101;  
Best Local Similarity 18.0%; Pred. No. 1.3e-11;  
Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

```

Qy 3 LWSPVLEVA-----CKETLOIDNRPRKEVMILKSKSNQPKNPKFPTVG 51
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 LSWYNSLHVADPVEAVLQLODCSIFIKIIDRIHGTGEOQILK-----QPVSERLDFVCSF 67
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 52 KTLQLKPTGAGTKMSVSWAAVQOKKRMGAAAAPPKOT-----WRYTAPSR-PAAWAHP 106
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 LQNRKHPS-----SPECIVSA-----QKVLGSELELAKWMTLLYHSTMSKSPDW--- 116
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 107 PPOAPLVKNVYKTPQBPVYI-----SPOD-----RSLKENLS 140
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 -----EQPEYKIQATLAVILKPVLDHEDGLNLEDLENFLQKAPVFCSTCSFPEELS 169
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 141 PMTPGNLLDVLDNLRFTPLNTRGKGQATFPDNLAAWP-----TPTLKGNVKSCANDM 194
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 170 P-P-PSH--QAKREIRFLELQKVASSSGNFLSSPSAPMGDILOTFQF--MRRLKKQL 223
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 195 RPRRITDDLEDOPATNKTFDVKHSETINTSLDLDCSRIDGQPHFTPLKTTTIVHATHT 254
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 224 ADERSNRDELELELAENRKLITKDAQIAMWQORID-----259
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 255 RALACIHEEGSPSP--PRT-----PTKSAIHLDKRDIKLVGSPLRKYSE 296
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 260 -RLALLNEKQAASPLEPKLEELBDKNESITMRJLHETLKQODLUKTEKSOMDRINKOLSE 318
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 297 SMKDLSSLSPQTKVAIOGSPNLENMKIRSTEQNRYQEQQIOIQLKADLNSSSSSEASLA 356
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 ENGDLSPKREFASHLQOLQDALNELTEEHSKATQEWLEKQAOLE-KELSAALQDKKCLE 377
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 357 GQOEFLFNHSEIILAOSSGRFNLHVGRKSVGVGS PVKNPHKRSHLSFSDAPSNSLYRNE 416
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 378 EKNEIL-----QGKLSQLEE-----HLJSLOQDNP-----401
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 417 TVAISPPKKORVEDTTLPRGAAPANASARSSSAHAWPHAOSSKFKLAOTMGLMKKPAFTR 476
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```









[illegible]

```

Query Match      2.6%; Score 250; DB 3; Length 2101;
Best Local Similarity 18.0%; Pred. No. 1.3e-11;
Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

QY      3 LVWSPVLEVA-----CKETQLIDNFRKEVMILKSKSNQPKNPKRKEPTVG 51
Db      12 LSWNSLHVADPVEAVLQDCSIFIKIDIRHGTEGQOILK-----QPVSERLDFVCSF 67

QY      52 KTLQLKPTGAGKTMKSVSAVQOKRMSAAAPPSKOT---MRVTAPSR-PAAWAHP 106
Db      68 LQKNRKHPs---SPCLVSA--QKVLGSELELAKMTLLLYHSTWSSKSPDW--- 116

QY      107 PPQAPLVEKNVYKTPQREPYI-----SPQP-----RSLKENLS 140
Db      117 -----EQFEYKIQBELAVILFVLDHEDGMLNDELNFLOKAVPFTCSSTFPPELS 169

QY      141 PMTPGNNLLDIVDLNRFTPLTETRGKQATIFPDNLAAP-----TPTLKGNVKSCANDM 194
Db      170 P--PSH--QAKREIRFLELQKVASSSSGNFNLSGSPASPGWDILQIPQFQ--MERLKKQL 223

QY      195 RPRRIPTDDELQDPATNKTFDVKHSSTINISLDTLDCSRIDQGPHTPLNKTTIVHATHT 254
Db      224 ADERSNDELELELAENRKLLTTEKDAQIAMMQORID----- 259

QY      255 RALACIHESEGSPF--PRT-----PTKSAIHLKRDIKLVGSPFKYSE 296
Db      260 -RLALLNEKQAASPLFPKLEELRDKNESITWELHETLKQCQDLKTEKSMQRKINQLSE 318

QY      297 SMKDLSLSPQTKYATQGSMPNINEMKIRSEIQNRYYEQEQIQIKADLNSSSSSSASIA 356
Db      319 ENGDLGSKLREFAHLQQLQDALNETEHSKATOEWLKQAOLE--KELSAALODKKCLE 377

```

Qy	357	GOEFLFNHSEILLAQS	RFLNHFVGRKSVKGS	PKVKNPHKGRSHSELS	SPDAPSNE	SLYRNE	416
Db	378	EKNEIL	-----QCKLSQLEE	-----	HLSQLQDNP	-----	401
Qy	417	TVAISPPKQORVEDT	TTLPRSAAPANASAR	SSAHAMPHAKSKFKLA	QOTWSLMKCKPATPR		476
Db	402	-----POEKGEVLGDVL	-----	-----	OLETLKQEAATL		425
Qy	477	KVRDTSIQPSVKLYD	SELYMOTCINCPDPA	ATTIDPFLASTWYLD	DSQAVDRHQADPKKW		536
Db	426	AANNTOQARVEMDET	ERGQEEA	-----KLAE	RGHFEE	-----KQO	463
Qy	537	INALVSTPADLDAN	NKIDVGKLFNEVRN	KELVVAPTKEEQSMNYLT	KYRLETLRKA	AAV	596
Db	464	LSSLIIT	-----DLQSSISN	-----LSQAKEE	-----	LEQASQAH	492
Qy	597	ELPFSEQMRLPCS	KVAVYVKNQALR	ITSRDNIHLDDVVMQ	BTILELLLCFNPLMLRIGLEV		656
Db	493	-----	-----GARLTAQV	ASITSEL	-----TTLNATI	-----	514
Qy	657	VFGKIQMOSNRDI	VGISTILNRLFNK	CCEQRYSKAYTL	-----TEEYARTIKKHS	LQKILF	715
Db	515	-----QOQDQELAGH	-----	-----KQAKEKQAO	LAQTLQOQEQASQ	GLRHQVQELSS	558
Qy	716	LLPFLDAQKRI	VGNPCILFVKKPS	PHKETDKILLRFS	SELLANIGDITRE	-----LRR	773
Db	559	SL	-----KQK	-----EQQLKEV	-----	AEKQEA	598
Qy	774	LQHRQTFLEDFY	AFNNL	AVDLRQDVGRL	TRVVEVILLRDDL	TQLRVP	832
Db	589	ABEREASLR	ERDAALQLEALEK	KAKL	-----EILQOOLQV	ANEAR	631
Qy	833	KLALGALGEAN	FOLGDDIAQD	IVDGHREKT	-----LSLLWQ	LIYKPSKPF	887
Db	632	-----	-----DSAQTSV	TOQREKAE	LSRKEVBELOAC	VEFARQOEHQAQVAE	675
Qy	888	KWRRHHLHVVI	QRRIRHKELMER	HRHAATVIOAVFR	GHOVRKVVKLP	FKTERTQAAITLOK	947
Db	676	-----	-----LELQ	RSQQA	KATEKVAQ	-----EKDQ	720
Qy	948	FTRYILAQKOLY	OSYHSIITI	QWRWAQOLGRO	HQRQRFVELRBA	AIFLORIMRRILF	1007
Db	721	EKRK	-----AADAL	BEQOQCISELKA	-----ETRSLIVE	QHKRKELEB	771
Qy	1008	LLAA	-----	-----AETARL	ORSQKQAA	ASTIQ	1052
Db	772	LQIGEAHQAE	TEVLRRELAB	AMAAQHTAB	SECBQLVKE	VAANR	824
Qy	1053	L	-----IMFVOR	MRKSMLEORKE	FOOLKRAANI	QORWRAKLSMRK	1109
Db	825	AQYGAMFQ	BQLMTLKEE	CEKARQELQ	AEKVGAGIESH	SELQISRO	875
Qy	1110	KVOA	-----YRKATI	QMRIDRNHY	SLRKVNI	CLQOELRAIMK	1168
Db	876	ELHANLARAL	QOQVEKEVRAQ	KLADDLSTLQ	EKWAATSKEVAR	LETMRKAGE	931
Qy	1169	YRMQOMI	QDRNAYLTRKCI	INVORRWATL	OMRRERKNYLH	LOTTTKRIQIK	1228
Db	932	TASRELIV	GPARGDR	-----	QPEWLE	EOQGRQ	976
Qy	1229	MKKORAEFLQ	KKVTLVQKRR	ALLQMKEREY	-----LHLREVT	IKLORFFHQA	1286
Db	977	MGNE	-----	-----LELRAAL	MESQOQO	EERGQERVARL	1022
Qy	1287	RAKYRGTQ	AAVSCLOMHW	RNHLIRKE	RNSFTQLROA	-----AITLQRRY	1343
Db	1023	-----EKAARAE	LEMLQNAL	-----NEQ	ORVEFATLQ	EALAHALTEK	1074
Qy	1344	YAQLQKAA	ITIQTRYP	AKAMQOVVY	OKREAI	IKVORRY	1395
Db	1075	-AQIKLEBLROT	-----	-----VKOLKE	QAKKEH	ASGSAE	1129

1396 ----EYKQKQAVIRLQKWRSSIRDMLCKA-----GYRRIRLS 1431  
 1130 SKLEQQCKQEQADSLB---RSLEARASAEARDSALETLQGLBEKAQELGHSQALA 1186  
 1432 SL-----SIQKWRATVOARRORIFLSTIRKVRMLQAFRATLILMRQQ- 1475  
 1187 SAQRELAARFKVKQDHSKADEWKAQVARGQ-----EAERKNSLISSLEBEVSILNRQV 1241  
 1476 -RRFEMKRAAVI-----QRRFRACMLKARODYQL--IOSVILVORKFRANR--- 1524  
 1242 LEKEGESKELRLVMAESKSKLESCACRQOPATVPPELQNAALLCGRCRASGREA 1301  
 1525 ----SMKQARQHFVQLRTIAVHL-----QKFRGKR-----LMIEQRNCFOLLRCSM 1567  
 1302 EKQVASENLQELTSQAERAEBELGOELKAWQKFFQKEQALSTLQLEHTSTQALVSELL 1361  
 1568 PG-----FQARARGFWARRFOALTPMMDLIROKRAKVIOYRWGVLIRRRKHQGLL 1623  
 1362 FAKHLCOQLOAQEAQAAEKHRE-----ELEQSKQAGGL-----RAELLRAQRELGELI 1410  
 1624 DIRKRIQLOEAKAVNSVRCKVOEAVRFLR 1654  
 1411 PLROKVAEQRTAQOLRAEKASYAEQSLMK 1441

RESULT 13

US-08-195-487-4  
 ; Sequence 4, Application US/08195487  
 ; Patent No. 5783403  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOUGRATLY, GARY  
 ; APPLICANT: LIDGARD, GRAHAM P  
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
 ; STREET: 53 STATE STREET  
 ; CITY: BOSTON  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US/08/195,487  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/901,701  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: PITCHER ESQ, EDMUND R  
 ; REGISTRATION NUMBER: 27,829  
 ; REFERENCE/DOCKET NUMBER: MTP-013  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/248-7000  
 ; TELEFAX: 617/248-7100  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2101 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-195-487-4

Query Match 2.6%; Score 248; DB 1; Length 2101;  
 Best Local Similarity 18.0%; Pred. No. 2e-11;  
 Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

Qy 3 LWNPSVLEVA-----CKETLQIDNRNFRKVMILKSKSNQPVKNPKPPTVG 51  
 Db 12 LSWNSLHVADPVEAVLQDQCSIFIKIIRIHGTEGQQILK-----QPVSERLDFVCSE 67  
 Qy 52 KTIQLKSPGTAGTKMKSVAANVQKKRMSAAAAAPPKOT-----WRVTASR-FAAAWHP 106  
 Db 68 LQKNRKHS-----SPECVLSA-----QKVLGSELELAKMTMLLYHSTMSKSPDM--- 116  
 Qy 107 PPQAPLVKKNVYKTFQEPVVI-----SPOP-----RSLEKNLS 140  
 Db 117 -----EQEYKIQABLAVILKEVLHDHEDGLNEDENFLQKAPVSPSCSTTFPELS 169  
 Qy 141 PMTPGNLLVDINLRFTPLTETRGKGOATIPPDMLAAMP-----TPTLKNVKSXCANDM 194  
 Db 170 P--PSH--QAKREIRFLELQKVASSSGNNFLSGSPASPMGDILQTPFQ--MRRLKKQL 223  
 Qy 195 RPRIITPDDLEDQATNKTFDVKHSETINISLDTLDCSRIDGQHTPLINKTTIVHATHT 254  
 Db 224 ADERSNDELELELAENRKULTTEKDAQAMQOORID----- 259  
 Qy 255 RALACIHEEGPSP--PRT-----PTKSAIHDLKRDIKLVGSPLRKYSE 296  
 Db 260 -RLALINEKQAASPLEPEKLEELDKNESLTMRLHETLKQCDLKTESQMDRKINQLSE 318  
 Qy 297 SMKDLSSLSPQTKYAIQGSMPNLMEMKIRSTEQNRYOEOQIQIKAKDLNSSSSSEASLA 356  
 Db 319 ENGLSPKLEFASHLQOLQDALNELTEHSKATQEWLEKQAQLE-KELSAALQDKKCLE 377  
 Qy 357 GQOEFLFNHSEILLAOSRFRNLHEVGRKSVKGPSVKVNPBKRSHELSPDASNESLYENE 416  
 Db 378 EKNEL-----QGKLSQLEE-----HLSQLODNP----- 401  
 Qy 417 TVATSPPKQORVEDTTLPRSAAPANASARSSAHAWPHAQSKFKLAQTMSLMKKPPR 476  
 Db 402 -----POEKGVLGDVL-----QLETLQEAATL 425  
 Qy 477 KVRTSTQPSVKLYDSLYMOTCINPDPPAATTIDFLASTWYLDQAVDRHQADPKW 536  
 Db 426 AANNTQLQARVEMLETERGQEA-----KLLAERGHFEE-----KQ 463  
 Qy 537 LNALVSPADLADANNKIDVGKLFNEVRNKLVAAPTKEQSMNYLTKYLETILRKAHV 596  
 Db 464 LSSLIT---DLQSSISN-----LSQAKER-----LEQASQAH- 492  
 Qy 597 ELFFSEQMRLPCSKVAVYVVKQALRIRSDRNHLHDVVMQRTILELLCFNPLWRLGLEV 656  
 Db 493 -----GALTAQVASLITSEL-----TTLNATI----- 514  
 Qy 657 VFGKIQMOSNRDIVGLSTFILNRLFRNKCEQRYSKAYTL-TBEYATIKKHSLOKTLF 715  
 Db 515 -----QQDQELAGL-----KQAAKEKQAQLAQTLQOQFQASQGLRHQVEQLSS 558  
 Qy 716 LLPFLDQAKOKRIYKHNPCLVKSKPHKETDILLRFSSELLAMIGDITRE--LRLRGV 773  
 Db 559 SL-----KQK-----EQOLKEV-----AEQKQATRODHAQOLATA 588  
 Qy 774 LQHRQTFLDERDYAFNNL-AVDLRDGVRLTRVVEVILLRDLTLRQLRVPALSRLOIFNV 832  
 Db 589 AEEREASLRERDAALKQLEALEKEKAAL-----EILQOQLOVANEAR----- 631  
 Qy 833 KLALGALGEANFQGGDIAADQIVDGHREKT-----LSLLMQLIYKFRSPFHAATVQL 887  
 Db 632 -----DSAQTSVTQAKREKAEKSRKVEELQACVETARQEQHEHAQVAE 675  
 Qy 888 KWRHRLHVLVVIORIRHKLMMRRHRAATVIAVFRGHOMKRYVKLFKTERTOAAIIQK 947  
 Db 676 -----LEQLRSEQKATEKERVQ-----EKDQLQEQALKEKLVKYGSLSE 720  
 Qy 948 FTTRYLAQKLYQSYHSIITIQRWRAQQLGRHQRFVLELREAAIFQRIWRRLFAKK 1007  
 Db 721 EKRR--AADALEEQRCISLKA--ETRSIVEQHKRERKELEE-----ERAGKGLAERL 771  
 Qy 1008 LIAA-----AETARLORSKQQAASVYIQ-----MQWRTYQLGRIQRIHFLRQD 1052



```
Db 772 LQLGEAQAETVLRRELAEMAAQHTAEBCEQOLVKEVAAWR-----DGVDSQOE 824
QY 1053 L----INFVORMRKSMLEFORKEFQOLRAAINIQQORWRKLSMRKCNADYALRSSVL 1109
Db 825 AQYGAMFQEQMTLKECEKARQOEAKVAGIESHSELQISRO-----QNKLA 875
QY 1110 KVQA-VKATQWIRDNHYSIRKKNVICLOQRRLAIMKMRQENYLRNASILVOKR 1168
Db 876 ELHANLALQVQOEKEVRAQKLADSLTQEKAAATSEKVALETIVKAGE-----QOE 931
QY 1169 YMRQOMIQDRNAYLTRKCIINQVRWRATLQMRERKNYLHLQTTTRIQIKFRAKE 1228
Db 932 TASRELVPKPARAGDR-----QPEWLEQGRQ-----FCSTQAALQAMEREAEQ 976
QY 1229 MKQRAEFLQKVTIVVQKRRALLQMRKEROEY--LHLREVTIKLQRRFHAKSMRFM 1286
Db 977 MGNE-----LERLAALMESQOQOEERQOEEREARLTOERGRAQADLAL- 1022
QY 1287 RAKYRGTOAAVSCLOMHWNRHLLRKRERNSTFLQROA---AITLQRRYRARNLMIKOLKS 1343
Db 1023 -----EKARAELERLQNAL--NEQVFEPTLQELAHALTEKSGKQELAKLGLA 1074
QY 1344 YALQKQAATITOTRYBAKXAMQOVVLYQKREAIIVQVRYRGNLE-----MRKQI 1395
Db 1075 -AQIKELBELROT---VKQLKEQLAKEKEHASGSGAQSEAAGRTEPTGPKLEALRAEV 1129
QY 1396 ----EYQKQRAVIRLOKWRWSIRDWLCKA-----GYRRIRLS 1431
Db 1130 SKLEQOQKQOQOADSLE---RSLEAERASRAERDSALETLOQLBEKAQELGHSQSALA 1186
QY 1432 SL-----SIOKWRATVQARROREIFLSTIRKVRMOAFIRATLLMQO- 1475
Db 1187 SAQRELAARTKVQDHSKAEDEKQAVARGQ-----BAERKNSLISSEBEEVSIINQV 1241
QY 1476 -REEFMKRAAVI-----QRRFRACAMLKARQDYQL--IQSSVILVQKFRFRANR--- 1524
Db 1242 LEKEGSKELKRLVMAESEKSKLEBSCACCRCQROPATVPPELQNAALLCGRRCRASGREA 1301
QY 1525 ----SMKQARQEFVQLRTIAVHL-----QKFRGKR-----LMIBQNCFOLLRCSM 1567
Db 1302 EKORVASENLQBELTSQABRAEELQOELKAWQEFKQEQALSTLQEHSTQALVSELL 1361
QY 1568 PG----FOARAGFEMARKEFOALMTPEMDLITROKRAAVIQRYWGYLIRROKHQGLL 1623
Db 1362 PAKHLCCQLOAQAAAEKHRE-----ELESGKAAGL-----PAELIRAQRELGELI 1410
QY 1624 DIRKRIAQRQBAKAVNSVRCKVQEAVERFLR 1654
Db 1411 PLRQKVAEQDRTAQQLRAEKASVYAEQLSMLK 1441
```

```
RESULT 14
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; Patent In Release #1.0, Version #1.25
```

```
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-06160-4

Query Match 2.6%; Score 248; DB 5; Length 2101;
Best Local Similarity 18.0%; Pred. No. 2e-11;
Matches 329; Conservative 275; Mismatches 647; Indels 580; Gaps 75;

QY 3 LVWSPVLEVA-----CKETLQILDNRNFRKEVMIILKSKSNQVKNPKRPFVTG 51
Db 12 LSWNSLHVADPVEAVLQLDQCSIFIKIIDRIHGTGTEGQOILK---QPVSERLDFVCSF 67
QY 52 KTLQLKSPGAGTKMKSUVSAVQOKKMSAAAAAPPSKOT---WRVTAPSR-PAWAHP 106
Db 68 LQNRKHPS---SPECLVSA---QKVLGSELELAKMTLLLYHSTMSKSPRDW--- 116
QY 107 PQQAPLVEKNVYTKPOEBPVYI-----SPOP-----RSLKENLS 140
Db 117 -----EGFEYKIQAEIAVLKFLVLDHEDGLNLEDLENFLQKAVPSTCSTTFEELS 169
QY 141 PMTPGNLLDVIDNLRPTPLTETRGQATIFPDNLAAMP-----TPTLKGNVKSANDM 194
Db 170 P--PSH--QAKREIRFLEQLKVASSSSGNFLSGSPASPMGDILQTPQFQ--MRRLKKQL 223
QY 195 RPRTYPPDLEQOPATNKTFFVKHSETINISLDTLDCSRIDGQPHPTPLNKTIVHATHT 254
Db 224 ADERSNRDELELELAENRKLTEKDAQIAMMOQRIID----- 259
QY 255 RALACIHEEGSPSP--PRT-----PTKSAIHDLKEDIKLVGSPLRKYS 296
Db 260 -RLALLNEKQASPLEPEKLEHLRDKNESLTWRLHETLKQCODLKTEKQOMRKINQLS 318
QY 297 SMKDLSSLSPQTKYAIQGSMPNLNEMKIRSIQNRYYQEQIQAIAKOLNSSSSSEASLA 356
Db 319 ENGDLSFKLREFASHLQQLQDALNELTEHSHKATQEWLEKQAQLE-KELSAALQDKKCLE 377
QY 357 GQOEFLFNHSEITLAQSSRFNHEVGKSVKSPVKNPKRRSHSELSFSDAPNESLYRNE 416
Db 378 EKNRIL-----QKLSQLEE-----HLSQLQDNP----- 401
QY 417 TVAISPKKKORVEDTTLPRSAAPANASARSSSAHAWPHAQSKFKLAQTMSLMKKPAETR 476
Db 402 -----POEKGVLGDVL-----QLETIKQBAATL 425
QY 477 KVRDTSIQPSVKLYDSELYMQTCINPDPPAATTTIDPFLASTMYLDBQAVDRHQADFKEW 536
Db 426 AANNTQOLQARVEMLETERGQEA-----KLLAERGHFEE-----KQ 463
QY 537 LNALVSIADLADLNNKIDVGLFNVEVRNKLNVVAPTKEBQSMYLYKRYLETURKAIV 596
Db 464 LSSLIT--DQSSISN-----LSQAKE-----LEQASQAH- 492
QY 597 ELFFSEQMLPCSKVAVYVKNQALIRSDRLHLVDVVMQRTILELLLCNPLNLRGLLEV 656
Db 493 -----GARLTAQVASTSEL-----TTLNATI----- 514
QY 657 VFGEKIQMSNRDVLGLSTFILNRLFRNKCEBQRYSKAYTL-TEEYAEITKHKSLQKIIF 715
Db 515 -----QQQDQELAGL-----KQAKEKQALQACTLQQOQKQASQGLHVEQLSS 558
```



Db 879 LGAHLGQHACADWQLSLRHAQAQSAVETLQ-----APLDSLREEQRLAEALEHLQ 932  
QY 1545 OXFERGKRLMIEORNCFOILRCSMPGFQARGFMARKRFOALMTPEMMDLIRCKRAKVI 1604  
Db 933 QORQ-----ROQDEFQRLQADWQAWRER-QDNLDDSRIDLALIG-----LSEEQATQ-- 977  
QY 1605 QRYWRGYLIRRRQK--HQGLLDIRKRIAQLRQEAQAVNSVRCKVQBAVRFLRGRFIASDA 1662  
Db 978 ---WRQLQRLQEBEITRQQTLEAERQAQLLOHRRQRPETDREALEDNLAQQORERLAASEQ 1034  
QY 1663 --LAVLSQL 1669  
Db 1035 AYLETYSQL 1043

Search completed: September 13, 2004, 17:51:32  
Job time : 49 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 17:43:10 ; Search time 192 Seconds  
(without alignments)  
3108.351 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVSPVLVEACKETQLI.....FISSVAFDTLCKLQIDMF 1861

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	672.5	7.1	898	US-10-108-260A-3399	Sequence 3399, Ap
2	408.5	4.3	726	US-10-188-832-82	Sequence 82, Appl
3	348	3.7	1022	US-10-437-963-156007	Sequence 156007,
4	300.5	3.2	1855	US-10-177-293-315	Sequence 315, App
5	291	3.1	1742	US-10-012-697-1548	Sequence 1548, Ap
6	284.5	3.0	1805	US-09-820-843A-73	Sequence 73, Appl
7	284.5	3.0	1805	US-10-282-122A-63513	Sequence 63513, A
8	277	2.9	1818	US-10-282-122A-64245	Sequence 64245, A
9	270	2.8	2383	US-10-082-830-260	Sequence 260, App
10	260	2.7	1879	US-10-205-823-419	Sequence 419, App
11	254.5	2.7	1837	US-10-369-493-22734	Sequence 22734, A
12	250.5	2.6	1151	US-10-408-765A-994	Sequence 994, App
13	250.5	2.6	1416	US-10-295-027-446	Sequence 446, App
14	244.5	2.6	1401	US-10-408-765A-2125	Sequence 2125, Ap
15	238.5	2.5	1893	US-10-408-765A-1696	Sequence 1696, Ap

16	238.5	2.5	3674	15	US-10-291-265-454	Sequence 454, App
17	238	2.5	1507	16	US-10-437-963-185157	Sequence 185157,
18	235.5	2.5	1750	14	US-10-128-714-8556	Sequence 8556, Ap
19	234.5	2.5	3899	14	US-10-171-311-4	Sequence 4, Appli
20	234.5	2.5	3917	14	US-10-171-311-8	Sequence 8, Appli
21	233.5	2.5	1965	15	US-10-369-493-3279	Sequence 3279, Ap
22	231.5	2.4	2295	16	US-10-437-963-135452	Sequence 135452,
23	230.5	2.4	1211	12	US-10-282-122A-66630	Sequence 66630, A
24	230.5	2.4	3907	14	US-10-171-311-2	Sequence 2, Appli
25	230.5	2.4	3935	14	US-10-171-311-6	Sequence 6, Appli
26	229	2.4	1601	16	US-10-437-963-185154	Sequence 185154,
27	228.5	2.4	1327	16	US-10-408-765A-1215	Sequence 1215, Ap
28	228.5	2.4	1679	15	US-10-369-493-22080	Sequence 22080, A
29	227	2.4	1707	14	US-10-128-714-3556	Sequence 3556, Ap
30	226.5	2.4	3911	15	US-10-370-685-100	Sequence 100, App
31	226.5	2.4	3911	16	US-10-408-765A-1839	Sequence 1839, Ap
32	226	2.4	3336	16	US-10-408-765A-2453	Sequence 2453, Ap
33	225.5	2.4	3225	16	US-10-408-765A-254	Sequence 254, App
34	225.5	2.4	5171	16	US-10-408-765A-2687	Sequence 2687, Ap
35	225	2.4	2052	16	US-10-437-963-137285	Sequence 137285,
36	225	2.4	3562	15	US-10-341-434-109	Sequence 109, App
37	224	2.4	1294	12	US-10-282-122A-61292	Sequence 61292, A
38	224	2.4	2020	15	US-10-369-493-5128	Sequence 5128, Ap
39	224	2.4	2020	15	US-10-369-493-5129	Sequence 5129, Ap
40	223.5	2.3	1827	15	US-10-369-493-5368	Sequence 5368, Ap
41	223.5	2.3	2310	9	US-09-874-923-120	Sequence 120, App
42	223.5	2.3	2310	9	US-09-991-496-120	Sequence 120, App
43	223.5	2.3	2310	10	US-09-820-843A-114	Sequence 114, App
44	222	2.3	1384	16	US-10-473-576-22	Sequence 22, Appl
45	221.5	2.3	2033	16	US-10-408-765A-419	Sequence 419, App

ALIGNMENTS

RESULT 1

US-10-108-260A-3399  
; Sequence 3399, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20040005560A1el full length cdna  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3399  
; LENGTH: 898  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-3399

Query Match	7.1%;	Score 672.5;	DB 15;	Length 898;
Best Local Similarity	25.2%;	Pred. No. 2.3e-44;		
Matches 236;	Conservative 182;	Mismatches 309;	Indels 209;	Gaps 30;
QY	850	IAAQIVDGH-----REKLSILWOLIIYKFRSP-----KFHAAATVLQKWRHML	895	
Db	43	IKLOAFVGVYVRKQMLQKRAVLSI--QSYFRMKARQYILKMYKAIIVIQNYI--HAY	98	
QY	896	HVVIQRRTHKELMRHRAATVIOAVFRGHQMKVKVLFKTRTQAAIILQKFTRYLAQ	955	
Db	99	KQVNVQR---KNFLQVKKAAATCLOAYRGYKVRQIK-----QQSIAALKIQSAFRGYNKR	151	
QY	956	KOLYQSVHSIIIFQGWRAQOLGRHORFVFLRRAAIFLQIRRRFLAKLLAAAEFA	1015	
Db	152	VKIQSVLQSIILKIQWRYAYKTLHTRTHFLTKAAVLSLQAYRGWKVKQIRREHQA	211	
QY	1016	-----BLORSQKQ-----QAAASYIQMWRTYQLGRIGRHEFLQRDLIMFVQ-----	1058	
Db	212	LKIQSAFRMAKQKQFLFKTAALVIQQNFRAWTAGRCQMEYIELRHAVALQVSNWKGK	271	



Db 664 ERI-----LYKQKNSSISIPETPTVTRIVSRKLPDVLRRDNMBEITNPLQAIOM 717  
 QY 1852 ILCKLQI 1858  
 Db 718 VMDTLGI 724

RESULT 3  
 US-10-437-963-156007  
 ; Sequence 156007, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yinhua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 156007  
 ; LENGTH: 1022  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_55716C.1.pap  
 US-10-437-963-156007

Query Match 3.7%; Score 348; DB 16; Length 1022;  
 Best Local Similarity 19.7%; Pred. No. 3.2e-18;  
 Matches 229; Conservative 188; Mismatches 441; Indels 302; Gaps 42;  
 QY 419 AISPCKQVETTLPRSAAPANASARSSAHAWHAQSKFKLAQTMGLMKKPAATPKV 478  
 Db 7 AVSSPFRD-LSNLTFRNPPRPNKSSASKEBPLP-----SATPTTR 47  
 QY 479 ROTSIPSVKLYDSELYMTCINPDPAATTTIDPFLASTMVLDEQ-----AV 526  
 Db 48 RRGPPPP-----PRGAATATPLARLRLELDQSRARRAESRGDAL 92  
 QY 527 DRHQADFKKWLNALVSIPA-----DLDA-DLANKIDV-----GKLFNE 563  
 Db 93 RAFASSATSWLSLLLRDPSACGAPSAARVTRDAPAHGVQGRDAVDGERARGSPKR 152  
 QY 564 VENKELVAPTKEEQSMNLYTKVRL-ETLRKA-----AVELFSEQMRLPCKSVAYV---- 614  
 Db 153 HRGEGDRGPGPRKRTTPAMAASURLSRLVCSLDVTRMGSHMSREACHEVLVMCO 212  
 QY 615 ----VNKQALRISDRNLHLDVVMQRTILELLCFNPLMLRLGLEVFG-----EKIQMO 665  
 Db 213 ICKNDINGLWKHECHPCPLVSDLRDLRDKATRIPCYNPKWLRIGLHVLGDSWIQNESQK 272  
 QY 666 SNRDVGLSTFLNRLF-----RNKCEQRYSKAYTLTEEXAETIKHSLQKILFL 716  
 Db 273 KQKEVAFLKVLKQLFVHMVTPSALNKAPEGHRRAC-----YAAASNSILKRLFL 326  
 QY 717 LPLDQAKOKR-----IVKENPCLFVKSPKHEKTDILLRPSSELLANIGDITRE 766  
 Db 327 VAALDRKATESGLPSESGIDLGGSPLFCRQTEIKSRQIVQESLGEVHMGEGDLIMH 386  
 QY 767 LRLGLVLOHRTFLDEPDYAFNNLAVLDGVRLEFRVVEVILLRDLRQLRV--PALS 824  
 Db 387 LNMVGLNYQALASEYDFTVGNLFEDQDGLICRIIQ--LLTSDASIIILKVIAPSDT 444  
 QY 825 RLQRIFNVKALGALGEANFOL-----GGDIAAQDIVDGHREKTLSLMLQIYKFRSPKPH 880

Db 445 YKRLHNCNTMAIQIKQAGFPLSDADGLSISAEIDVNGDKELLALLMNNFIYQPLVLV 504  
 QY 881 AAATVLQKWRHHLHVVIQRRIRHKELMRRAAAVIOAVFRGHOMRKVKVLFKTRTQ 940  
 Db 505 NETSVAQE-----ISRLKAPVSEQSI-----SEMKS 530  
 QY 941 AAILIKQFTERRVLA--OKOLYOSYHS-IITIQWWEAQQLGRQHRQFVELREAAIFLQ 997  
 Db 531 QTGLLYDWIQETSLDCRKEFLSCHTDMADIITYOFNNGIKVLAQ-----FLQD 580  
 QY 998 IWRRRLFAKKLAAAEARLQSRQKQAAAASYIOMQRTYOLGRIQHEFLR----- 1049  
 Db 581 LPGWNLANDVLFVEKSA-----IILAFSLSHLTNVERLQKLNILDSKLD 627  
 QY 1050 QRDCLTMFVQRRMRKWSMLBQREKFPOLKR-----AANIQORWRAKLSMRK--- 1096  
 Db 628 HQSLVTEVSPRRSR-GTTDMKCHFPQTEETDGSRTREWAATVIQTOARRLNAMSKYCK 686  
 QY 1097 -----CNADYLALRSSLVKQVAYRKATIQMIDRNHYSLRKNVICLQORLAIM--K 1147  
 Db 687 LKNATQPCNKGHDPASSSPKSIADSSC-----IDSATKLVCEDDVDCSSNCQVLFYHD 741  
 QY 1148 MREORENYLRLRNASTL--VQKRYMRQOMIQDRNAYLRTRKCIINVQRWRATLQMRRE 1205  
 Db 742 PVSTKVDFLCRKAWAARKIQFAYRRFAHRIRSRIS-----AAIKIQSHWRC----- 788  
 QY 1206 RKNYLHLQTTTIRIQIKFRAKEMKQRAEFQOLKVTILVQKRRALLQMKRQREYLH 1265  
 Db 789 -----FSVRIFRKQI--QNTITTIQAVARLLVTGIWKRKSVIVIAHVRGWI- 832  
 QY 1266 LREVTIKLQRRFHAQSMFRMAKYRGTOAAVSCIQMHWNLRLKRKRNFLQRLQAAI 1325  
 Db 833 ARQTAVRNKKR-----ITTIQSYVKAYLLRLRKSQ-----EITDDIM 869  
 QY 1326 TLQRYRFLANMIKOLKSYAOLKQAAITITQYR-----AKKAMQKQVLYQ 1372  
 Db 870 CLINRLIAAVSQ-----RSISTIRQICATLSTATEHSEKCCQITIVNAGAVEILLKQINLEN 925  
 QY 1373 K--OREALIKVORRYRGNLEMRKQIEVVQKORQAV-IRLOKWRWSIRD-----MR 1419  
 Db 926 RGVPDQVILKQVLFTRNIARFNLQPVLANTPQAVEIVFQBLRLSKTEGFFVACDILKR 985  
 QY 1420 LCKA--GYRRIRLSSLSIOR 1437  
 Db 986 LCSEBEGHKLARALKRHIR 1005

RESULT 4  
 US-10-177-293-315  
 ; Sequence 315, Application US/10177293  
 ; Publication No. US20030124128A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Glatt, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Gannavarpu, Manjula  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Mertens, Maureen  
 ; APPLICANT: Myer, Vic  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: McNahan, John  
 ; APPLICANT: Meyers, Rachel E.  
 ; APPLICANT: Bast Jr., Robert C.  
 ; APPLICANT: Hortobagyi, Gabriel N.  
 ; APPLICANT: Pusztai, Lajos  
 ; APPLICANT: Meric, Funda  
 ; APPLICANT: Sahin, Aysegul  
 ; APPLICANT: Mills, Gordon B.  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
 ; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-038





APPLICANT: Leshkowitz, Dena  
 APPLICANT: Kita, David  
 APPLICANT: Garcia, Veronica  
 APPLICANT: Jones, Lee William  
 APPLICANT: Stache-Crain, Birgit  
 TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS  
 TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE  
 FILE REFERENCE: 2300-16252  
 CURRENT APPLICATION NUMBER: US/10/012,697  
 CURRENT FILING DATE: 2003-01-21  
 PRIOR APPLICATION NUMBER: 60/254,648  
 PRIOR FILING DATE: 2000-12-07  
 PRIOR APPLICATION NUMBER: 60/275,668  
 PRIOR FILING DATE: 2001-03-13  
 NUMBER OF SEQ ID NOS: 1568  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1548

LENGTH: 1742  
 TYPE: PRT  
 ORGANISM: Homo sapiens

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

FEATURE:  
 NAME/KEY: VARIANT

LOCATION: 1015  
 OTHER INFORMATION: Xaa = Any Amino Acid

779 TFLDSEFYAFNNLAVDLRDGRLTR-----VVEVILLRDDLTQLRVPASLQOR 828  
 691 TTFEYS-----RYGILMTQBELSFSKVECKVVLHR----- 723  
 829 IFNVKIALGALGEAN-FOLGGDIAAQDIVDGHREKTLISLWLIY--KFRSPKFAAAATV 885  
 724 -----LIQDSNOYQFG-----KTKIFFRAGOVAYLEKLRLDKLQSCVM 762  
 886 LQWWRHHLVVIQRRIRHKLMMRHRRAATVIAVFRGHQ-MRKVYKLFKTERTOAAII 944  
 763 VQK-HMRGWL-----QRKKFLRRAALITQYFRGQQTVRKAITAVALKEAWAAII 813  
 945 LQKFTFRYLAQKOLQSYH-SIITQWRRA-----QQLGRQH-----R 982  
 814 IQHCKRGYLV-RSLYQILRMATITMQYSGRFLARRRYRMLEEHKAVILQYKARAWLAR 872  
 983 QRFVELREAAIFLQRIWRRLFAKLI-----LAAATARLQSORQQA 1026  
 873 RRFQSIIRFVLNIQTYRVQRIQKLEQDKENHGLVEKLTSLAALRAGDVEKIQKLEAE 932  
 1027 ASYIQMWRTYQ-----LGRIOH--BFLRQDLIMFVORMRKSWMLEOR 1071  
 933 LEKAAATHRENYEKGKRYRDAVEEKLAKLQKHNSELETKQEI--QLKQKTEBELKEK 989  
 1072 KEFQQLKRAAINIQORWRAKLSMRKCNADYLAIRSSVLKQAY-----R 1115  
 990 MD-NUTKQLFDDVQKEEROMLEK-----SKELATQYEQIQLKEIKALKDE 1039  
 1116 KATIOMRIDRNHYSS--LFKNVICLQORLAIMKREORENYLRNASILVQK-----R 1168  
 1040 KMQLQHLVEGEHVTSDGLKAEVARLSKVTKTISEPEKEIE--LIQAQKIDVEKHVSQK 1096  
 1169 YMRQOM-----IQDRNAYLPRKCIINVQ-----RMRATLQMRERKNYL 1210  
 1097 ROMREKMSBITKQLLESYDIED---VRSLSVEDLEHNLGELMFA-----YE 1142  
 1211 HLQTTTKRIQIKFRKREMKQORAEFLQKLVTVVQKERRALLQMRKEROEVLHREVT 1270  
 1143 GLKKAIVLSEHSPQKQCYEKEIEALNPKVVHL-----SQTINHLOKL- 1186  
 1271 IKLQRRFHAQKSMRFRKAYRGTAQAVSCLQWHRNHLKRENSFLQROAAITLQRR 1330  
 1187 --FRENDINESIRHEVTLTSENMMIPDFKQOISE--LEKQKQDLEIRLNEQA-----EK 1238  
 1331 YRANLMIKQLSKYLAQLKQAATITQRYRAKAKAMQVVLVYOKQREAI--KVORRYRGNL 1389  
 1239 MKGKLEELSNQLHRSQEEGT-----QRKALBAQNEIHTKEKEKLIDKIQEMQEASD 1290  
 1390 EMRKQIEVYQKQ---QAVIRLQKWRMSIR--DM--PLCKAGYRIRLSLSI----- 1435  
 1291 HLKQFETESVKCNFRQASRLTENRDLBELDMKORVKKLQDQVTKLSTKTIKAND 1350  
 1436 -----QRKWRATVQARRQREI-----FLSTI 1456  
 1351 VHSSSGPKYELGMLQYKREDEAKLQNLILDLKPRGVVNMIPGLPAHLFWCVRVADSL 1410  
 1457 RKVRLMQAIPATL-----LMRQRRREFEMKRAAVIQRFRARCAMLKARQYOLIOS 1511  
 1411 NDANMLKSLMNSTINGIKQVVKHEHLEDFEM-----LSFWLSNTCHFLNCLKQYS--- 1459  
 1512 SVILVQRFRANRSMKQAR-----QEFVQ--LRTIAVHLQKFRKRLMTIQRNCFL 1562  
 1460 -----GEEEPFKHNSPQONKNCNLPDLSEYRQILSDVAIRIYHQF-----IIMEKNIOPI 1511  
 1563 LRCSMPGFOA-----RARGFWARKRFOALMTPF---MMDLIRQKR---AAKVIQRYMRG 1610  
 1512 IVPGLMYESLOGISGLKPTGF--RKRSSIDDGTGWTMSVLQLSYFYTTMCQGLDP 1569  
 1611 YLIRRRQKH-----QGLLDIKRRIAQLRQEAKANVSVECKVQEAVERFLRGR-----F 1657  
 1570 ELVRQAVQLFLIGAVTINSFLAKCMSCRGKMQ-----IRCNISYLEBWLKDLKQLNS 1625

Query Match 3.1%; Score 291; DB 15; Length 1742;  
 Best Local Similarity 17.8%; Pred. No. 2.9e-13;  
 Matches 292; Conservative 275; Mismatches 513; Indels 560; Gaps 77;

QY 415 NETVAISPPKKORVEDTTLPRSAAPANASARSSAHAPHAQSKKPKLAQTWSLMKKPAT 474  
 DB 382 SETV-----VKPMTFRQAVNARDALAKKIYAHLPDFIVE---RINQALQFSQKQHT 429  
 QY 475 PRKVRDTSIQPSVKLYDSELYMOTCIN----- 501  
 DB 430 FTGLVDIY---GFETFDVNSFFQFCINYANEKIQOQFNHVFKEQSEYMKEDIPWTLID 486  
 QY 502 -PDPPAATTTIDPFLLASTWYLDQAVDRHQADPKKMLNALVSPADLDADLNKIDVGKL 560  
 DB 487 FYDNQPVVIDIEAKGILELLEDECLLPHTGD-ENWLQKL-----YNNFYNRNPL 535  
 QY 561 FNEVR--NKLVV---APTKEQSMNLYTKYR-----LETLRKAARV-----PFSEQMR 605  
 DB 536 FEKPRMSNTSFVIOHPADVEYKCEGFEKRNRTVYDMLVEILRASKFHLCANFFQENPT 595  
 QY 606 LPCSKVAVVYNKQALRIRSDRNHLHDVVMQRTLELLCFNPLWLELGLVEVGEKIQMQ 665  
 DB 596 PP-----SP-----FGSMITVK 607  
 QY 666 SNRDIVGLSTFILNRLFRNKCBEORYSKAYTL-----TEEYAEITKKHSLQKILFILP 718  
 DB 608 SAKQVILKN---SKHFRITVSGKFSYLLMETLNATPHVRCIKNDEK-----LP 658  
 QY 719 FLDQAKQKRIKHNPCFLFKVKSHPKTKDILRFSSELANTIGDITRELRLGLVYLQHRQ 778  
 DB 659 F--EFDKRIQVQLRACGVLET-----IRISAQSYPS-----RW 690



## RESULT 7

US-10-282-122A-63513  
/ Sequence 63513, Application US/10282122A  
/ Publication No. US20040029129A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Wang, Liangsu  
/ APPLICANT: Zamudio, Carlos  
/ APPLICANT: Malone, Cheryl  
/ APPLICANT: Haselbeck, Robert  
/ APPLICANT: Ohlsen, Kari  
/ APPLICANT: Zyskind, Judith  
/ APPLICANT: Wall, Daniel  
/ APPLICANT: Trawick, John  
/ APPLICANT: Carr, Grant  
/ APPLICANT: Yamamoto, Robert  
/ APPLICANT: Forsythe, R.  
/ APPLICANT: Xu, H.  
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
/ FILE REFERENCE: ELITRA.034A  
/ CURRENT APPLICATION NUMBER: US/10/282,122A  
/ CURRENT FILING DATE: 2003-02-20  
/ PRIOR APPLICATION NUMBER: 60/191,078  
/ PRIOR FILING DATE: 2000-03-21  
/ PRIOR APPLICATION NUMBER: 60/206,848  
/ PRIOR FILING DATE: 2000-05-23  
/ PRIOR APPLICATION NUMBER: 60/207,727  
/ PRIOR FILING DATE: 2000-05-26  
/ PRIOR APPLICATION NUMBER: 60/230,335  
/ PRIOR FILING DATE: 2000-09-06  
/ PRIOR APPLICATION NUMBER: 60/230,347  
/ PRIOR FILING DATE: 2000-09-09  
/ PRIOR APPLICATION NUMBER: 60/242,578  
/ PRIOR FILING DATE: 2000-10-23  
/ PRIOR APPLICATION NUMBER: 60/253,625  
/ PRIOR FILING DATE: 2000-11-27  
/ PRIOR APPLICATION NUMBER: 60/257,931  
/ PRIOR FILING DATE: 2000-12-22  
/ PRIOR APPLICATION NUMBER: 60/267,636  
/ PRIOR FILING DATE: 2001-02-09  
/ PRIOR APPLICATION NUMBER: 60/269,308  
/ PRIOR FILING DATE: 2001-02-16  
/ Remaining Prior Application data removed - See File Wrapper or PALM.  
/ NUMBER OF SEQ ID NOS: 78614  
/ SOFTWARE: PatentIn version 3.1  
/ SEQ ID NO 63513  
/ LENGTH: 1805  
/ TYPE: PRT  
/ ORGANISM: Mycoplasma genitalium  
US-10-282-122A-63513

Query Match 3.0%; Score 284.5; DB 12; Length 1805;  
Best Local Similarity 19.2%; Pred. No. 1e-12;  
Matches 310; Conservative 264; Mismatches 611; Indels 433; Gaps 75;

QY 296 ESMKDLSSLPOTKY-----AIGSMENLNMKIRSIQNRYYQ-----EQOIQK 341  
DB 208 ENQKRLSLEYNTRELYSADNELQNVENIDQOIQFKHQYQYTRDBLSQLERKIQLT 267  
QY 342 AKDLNSSSSSEASLACQGBFLFNHSEILAQSRFRNLFHEVGRSVKSGSPVKNPHKRSHL 401  
DB 268 KQELVDKESALRVKIDDAFYINAR--LAE-----LDDVAK-----QL 303  
QY 402 SFSD--APNESLYRNETVAISPPK-----KORVEDTTLPRSA-APANASARSSSAHAWPH 454  
DB 304 SFQDGTITKQNAQVEDKLVANKEKDRLNTQKEAFNLRQASALIDINKLQENELFA-KH 362  
QY 455 A--QSKKEFLAQTMSLMK---KPAFPRKVRDTSIQPSVKLYDSLYMTCINPDPFAAT 508  
DB 363 LEHQNEFFQKQSDSLKULETEYKALQHKINEFNSATK--SEELLNQ---ERELEFKR 417  
QY 509 TTIDPFL--ASTMYLDE---QAVDRHQADFCKWLNALVSIPADLDADLN-----NKIDV 557

DB 418 REIDTTLTQASLEYEHQRESSQLLKDKQNEVKQHFQNLVYAKKELDKERNLLDQKKVDS 477  
QY 558 GKLFNEVRNKLWVAPTKEEQSMNYLTKYRLETFLRKAARVLPFSEQMR-----LPC 608  
DB 478 EAIF---QLKEKVAQERKELEL-YLVKQKQDKQKENEL-LPFEKQLKQHQADFENELEA 532  
QY 609 SKVAVVVKOALRIRS-----DNLHLDVVMORTILELLLCFNPLWLRGLGVVFGKIQ 663  
DB 533 KOELFEAKHALE-RSFIKLEDKKDLNTKAQO-----IANEFSQ 571  
QY 664 MOSNRDIVGLSTFILNRLFRN-KCEQRYSKAYTLTEEYATIKKHSLOKILFLPLDQ 722  
DB 572 LKTDKSKSADFELMLQNEVENLQEKQLFOERTYFERNAAVLNRLQKRELL-----626  
QY 723 AKQKRVKRNPCLVFKKSPHKETKDILLR-PSSELLAN-----IGDTRELRRLRGYVL 774  
DB 627 -----QOKETLDQLTQSFQERLINQREHKLVALSVKQKEILGKGL 668  
QY 775 Q-HROTFLDEFDYAFNNLA---VDLMDGVRLTVVVEVILLRDDLTQLRVPALSRLOIF 830  
DB 669 QDFSQTSLN-----ASKNLAEEMAIKFEKEIEATEKQLLND-----VN 708  
QY 831 NVKLALGALGEANFOLGSD-----TAAQDIVDGHREKTLTSLWLQLIYKFRSPKFAAAATV 885  
DB 709 NAEVIOADLAQLNQLNQRSELSQNAKQRIADPHNDSLKKL-----749  
QY 886 LQKWRERHVLHVIOQRIHKEIM-----RRHRAATVIOAFRGHQRKVKVYLPKTBRTQA 941  
DB 750 -----NEYELSLOKRLQLOELQTLQANQKHSYQN-QAYFEG-----ELDKLNREKQ 793  
QY 942 AI--ILOKETRRYLAQOLYQSVHSITTIQWRWRAOQLGROHQRQFVELREAIQLRIW 999  
DB 794 AFLNLAKKQTMEDVDAIKORLSKQKQALNNQC-----AELDRKTH-----ELNNA--FLNHDA 843  
QY 1000 RRLFAKCLLAAAEATAR---LQRS---QKQAAASYI---OMQW-----RTYQLGRI 1042  
DB 844 DQKSLQDQLATVKEQTKLIDLSALLERQFAENVAGFKRHSNKTSLQKQIYELTKK 903  
QY 1043 QRHEFLRQDRLIMFVORRMRSKWMLEQRK--EFQOLKAAINIQRWRAKLSMRKCNAD 1100  
DB 904 QSEQOTQKETELKIAFSDIQDYQVPELOQDQEFRO-----IEAKQRELDKLA-EKNNQV 957  
QY 1101 YLALRSSVLKVOAYRKATIQMID-RNHVYSRKKNVICIQLRLALMKWREQ-----1151  
DB 958 KLELNRFOALQKQDQTVQAOLELREHQLNLEQTAQFQANESLQKREQLTKKIQAF 1017  
QY 1152 -----RENYLRNLNASIIQKRYRMQMIQDRNAYLRTKCIINVORRWRAQLQWRE 1205  
DB 1018 HVELKKRNQFLAKGKRLFAKQDQQRKD-QEIN-----WR-----FKQF 1056  
QY 1206 RKNYLHLQTTYR-----IQIKFRAKREMKQKQAEFLQKLVKTVVQKERRAL 1253  
DB 1057 EKEYTDFDAKRELEBELEKIRRSLSQSNVELERKREKLATDTNLNKVQHNQTINRD--1114  
QY 1254 LQWKRERQBYLHLREVTIKLQRRFHAOKSMRFR-----AKYGTQAAVSCLOMHWRN 1306  
DB 1115 -QNSQIROFLERKNFQFSEANAKGAFILKRLSPASNLKIQEALAIQKLEPDKD 1173  
QY 1307 HLKRRERSFQLQRAAIIQLRRYRARNLMIKQKSAVLAQKAAITTIQTRYR-----AKK 1362  
DB 1174 EQQKELQOATQLQBFQFEKQ-----NPDIEKQ-----QLVAIKTQCEKLSDEKK 1220  
QY 1363 AMQKQVLYQKQREAI-----KVQRRYRGNLEMRKQIE-----VYQKQRAVI 1406  
DB 1221 ALNOKLVELKNLSQTYLANKNKAESQOQLOQKYTNLLDLKENLERTKQDQKHSIEFA 1280  
QY 1407 RLOKQWRSIR-----DMRLCKAGYRIRLSLSLQKWEATV-----1443  
DB 1281 RUTKFANDURFEKKQLLKAQRIVDKRLKENERNLHPLSNTERK-RAVLEQDISYPE 1339  
QY 1444 -QARRQREIFLSTIRKVR-----LMQAFIRATLL-----MRQQRREBFEMKRAAV 1487

Db 1340 KORKQATDAILASHKEVKKEGELQKLLVELETRKTLNNDFAKFSQRBEFENQRLKLL 1399  
 Qy 1488 VIQRFRARC-----AMLKARQDYQLIQSSVILVQKFRANRS-----MKQARQBFV 1534  
 Db 1400 ELQKTLQTNNSNFKTKAIQEIENSYKRGMEELNFKQKPEFDKNSRLYEFYRQORDEIE 1459  
 Qy 1535 QLRT-IAVHLOQKPRGRKRLMEQRCNCFQLLRCSMPGFQARARGF-----MARK 1581  
 Db 1460 RKESQVKLVKETOQRKANLLEAQAANKLNIEKNTTDFKEKELKAFKDKVDQDIDSTNKQK 1519  
 Qy 1582 RFQALMTPEWM---DLIRQKAA-----KVTQYVWVGVLIRRRKHQGLLDIRKRIA 1630  
 Db 1520 ELNBLNENKLLQSLIERERAINSKDSLNNKTIETI-----KROLH-----DKEMRVL 1568  
 Qy 1631 QLRQBAKAVNSVRCVQEAVERFAGRFIAPDALAV-----LSQDLRLSRIVPHLLMW 1682  
 Db 1569 RLVDRMKLABQ---KYOTEINRLRTQTFDSEKQDKIKNFPPLFLKINGNDMAFPYLYPW 1623

RESULT 8

US-10-282-122A-64245  
 ; Sequence 64245, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 64245  
 ; LENGTH: 1818  
 ; TYPE: PRT  
 ; ORGANISM: Mycoplasma pneumoniae  
 ; US-10-282-122A-64245

Query Match 2.9%; Score 277; DB 12; Length 1818;  
 Best Local Similarity 18.0%; Pred No. 4,1e-12;  
 Matches 311; Conservative 284; Mismatches 596; Indels 536; Gaps 75;

Qy 273 TKSAIHDLKRDIKL-----VGSPLRKYS-----SMKDLISLSPQTKYAIQGSMPNLNE 321  
 Db 272 TKQELVDEBETLKVRLNDADFYINSRLAELDLTSKINERDFVSKEQADVKASLANLTK 331  
 Qy 322 MKTR-SIEQNRVYQBOQIQIKAKDLNSSSSSEASLA---GQOEFNFHSEILAQSSPNL 377  
 Db 332 EKERLSAEKDSF--ERLUNTALNDINRMEQENALFAKHEQQOYEFRRKQ---QESLLKL 386  
 Qy 378 HEVGRKSVKSPVKNPKHRSHELFSFSDAPSNESLYRNETVAISPCKKQKQVEDTTLPRSA 437  
 Db 387 -ETHKQL-----QKRIGEFKIESEAKSEALLIQERELL---EKREIDILLTQAS 433  
 Qy 438 APANASARSSSAHAWPH-----AQSKTKFLAQTMSLMKKPA 473  
 Db 434 LEYEQORRTNQVLKEKHQVQOHPONLVHAKKLDQKRYLABQKRIDEEQIFKLKEKIA 493  
 Qy 474 TPKVRDTSTQPSVKLYDSELYMQTCINPDPFAATTIDPPLASTMWLDEQAVDRHQAD- 532  
 Db 494 TERRELE-----KLY-----LVKKQKQDKQKNDL 517  
 Qy 533 --PKWLNALVSI PADLDADLNKKIDVGKLFNEVRNKKELVVAPTKERQSMNLYTKYLET 590  
 Db 518 LIFEKQLR-----QYQADFENEI-----EEKQNELFASQKSLOKSFQLNKKAEL 563  
 Qy 591 LRKAAVELFFSEQMRLPCSKVAVYVVKQALRISDRNLHLDVVMQRTILELLCFNPLWL 650  
 Db 564 NOKA-----QKIA-----EDWAHLKQKHHAD-----L 587  
 Qy 651 RLGLEVVFGKIQMSNRDIVGLSTF-----ILNRLFRNKCE-----QRYSKAYTL 697  
 Db 588 EIFLEGEFNH-LQOEKHKLLLEARTQFQDNVSLLSARFQKQAEVLVQKOSLEQUTAFPNK 646  
 Qy 698 TEYAETIKKHSLOKILFLPFLDQAKQKRIKVNPCFLFVKKSPHKETK-----DILLRFS 753  
 Db 647 EQEAVERDWDRLANL-----EQKEMGLGDKVHQFDENSLNISKLAERELAIKFK 697  
 Qy 754 SELLA-----NIGDITRELRRLGVVQ-----HRTFIDDEFYAFNNLAVD 794  
 Db 698 EKELEAAQKQLSLDNNNNAGLKQLDKLSLKTERTLELEASKERILDFYDESSRIA-D 756  
 Qy 795 LRQGVRLTRVVEVILLRDDLTQLRVPALISLQRIFNVLKALGALGANFOLGGDIAAQD 854  
 Db 757 YESDLQ-ARLAEVKTLEKN---QOETAASERE-----LKVALEKLNQA-----796  
 Qy 855 IVDGHRKRTLSLWLQIYKFRSPKFHAAATVLQKWRHRHLHVVIQRRIRKELMRHRA 914  
 Db 797 -----KKAFLQIRKQOLLEIASVK-----QQLAKKANLLKNQQA 830  
 Qy 915 ATVIQAVFRGHQMKYKVLKFTERTQAAIILQKFTRYLAOKLYQSVHSIITQWRRA 974  
 Db 831 -----ELDK-----QTELEAAFLQDQTDK-----KELEKALHSV-----860  
 Qy 975 QQLGRQHRQREVELREAAIFLQIRWRRRLFAKLLAAAEATARLQRSQKQAAAAYIOMOW 1034  
 Db 861 -----KSKQELLE--RERSFLQK-----QREFAHVAGFRQVHF 894  
 Qy 1035 RYQQLGRI-----QRHEFLRQDRL-IMFYQRRMRSKWSMLBQKKEFPQQLKRAAINIQ 1086  
 Db 895 KTTQMRLSHPNKQOQSQEQIKRETELKTAFAADLKQYQLFELQKNQBFQ-----IEQ 947  
 Qy 1087 RWRKLSMRKCNADYLALRSVLKQAVRKA-----TIQWRID-----RNHYSLRKN 1134  
 Db 948 KHK-----ELELLAQKAEKQLEKQKATALASQDQDVQAKDLARQOHELELRON 999  
 Qy 1135 VICIQQRLRAIMKMRQORENVLRLNASILVQKRYR--MEQOMIDORNAYLRTKCLIN 1191  
 Db 1000 --AFNQASLSLNKQOREQULTNQVKVHGL--KKRHEKLTIKDRLIAEKEDQHKDAEIN 1055  
 Qy 1192 VQRWRATLQMRERRKNYLHQTTRK-IQIKFAKREMKKQRA-----EFTQL 1239  
 Db 1056 -QR-----PKQFENEYADPDQAKKRELQELQINRLNLEQSNASLLKRNQLTLDFAILL 1107  
 Qy 1240 KKVTLVQKRRRALLQMKERQEVYHLREVTIKLQRRFHAQKSMFMEAKYGTQ-----1294

```

Db 1108 KVVQHTQNR---VQMTQIQEFLLEKKNFQKASDEAALQKALLIKRLRSFASKLQQR 1164
QY 1295 --AAVSCIQMWRNHLKRRKNSFLQURQAATITQRRYRARNIMIKQKSYAQ----- 1346
Db 1165 EALAIQKLEFDRDBQKSEINNAKIQLEFQKLEKQNFDEAKQKQIEFQDQCORLDVEK 1224
QY 1347 --LKQAATIQRYRAKKAMQKVVLVYOKOREALIKVQRYRGNLEMRKQIE- 1396
Db 1225 RLLKQKLVOLKMLSKSYLYTKNRADLSQOO-----LQHKYANLJELKEKQKQAKRALDK 1278
QY 1397 ----VYQQRQAVIRLQKWRISIRDMRLCKAGYRIRLSSLSIQR-----KWRATVQARRQ 1448
Db 1279 KHRATYKMAQFVSEL-----RQEKQLLSAQKQVDDKSLLEQNQRH 1321
QY 1449 REIFLSTTRKVLMOAFIRATLLMRQORREFPMKRAAVIQRFRFARCAMLKARQDYOL 1508
Db 1322 LQNLSSSTKKRQS-----LEHDINKFDQRRKEAV----- 1351
QY 1509 IOSSVILVQKFRANRSMKQARQEFVQLTIAVHLQOKFRGKRLMLEQRNCFOLLRCMSP 1568
Db 1352 ----SSIL-----NSHKKLKQKQEGELQIL-----OKLSLAKTQIEQ--FSKL----- 1389
QY 1569 GFQARARGFMARKRFQALMTPEMDLIQKRAAKVIQRYWRGYLIRRRQKHQGLLDI--- 1625
Db 1390 -YQOREK-----LDQRTTLKHLRELKQNEATAHKNEVLEIENY 1430
QY 1626 -RKRTAQLRQBAKAVNSVCKVQEAUFURGRPIASDA--LAVLSQDLRLSRVPHLLMW 1682
Db 1431 YKKEQLRLTTEKSEFDNNKNRLFEYFKIRNEIEKKAHIKTVLETKKR-----HL--- 1483
QY 1683 CSEFMSFTCYGIMQAIRSEVDKQIL-----ERCSRIILNARYNSTVTNTVFOE 1731
Db 1484 -----VETRAVKHLQKQSIISKGQELKEIKERVSRDISHTNKQRELSLHQ 1532
QY 1732 GGLV--TIAQMLLRCKDSIEFNTLCTLIWFAHCPKRIKH--DYMTNPEAIYMWRE 1787
Db 1533 NKLLQKNLAEREREINNKDSLITQKIQTAKQLSE--KEARILKLEKMRVAVEQQYQA-E 1589
QY 1788 TKKLVARKEKMQNARKP--PMTSGRYKSQKINFTPCSLSPLEPDF 1832
Db 1590 ITRLKTRNADLEKNDKHLFPPLF-----KINGNDMNPYPYPWF 1629

```

## RESULT 9

```

US-10-082-830-260
; Sequence 260, Application US/10082830
; Publication No. US2003007604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 260
; LENGTH: 2383
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-260

```

## Query Match

Best Local Similarity 2.8%; Score 270; DB 14; Length 2383;

Matches 337; Conservative 306; Mismatches 605; Indels 512; Gaps 79;

```

QY 223 NISLDTDCSRIDGPHPTLNKTTTIVHATHTRALACIHERBGPSPRPTTKSAIHLDKR 282
Db 343 SLELDSSIFSQPDYQ---DADKALILVRSLVTR-----RQAVQDLRQ 382
QY 283 DI-----KLVSPLRKYSEBMK-DLSLLSPQTKYAIQSGMNPNNEMKI 324
Db 383 QLACQCEAVNLQQOHDQWEEBEGKALRQLKLTGERDTLAGQT-VDLQGEVDSLSKER- 440
QY 325 RSIEQNRYYQEOQIQI-----KAKDLN-----SSSSSEASLAGOQEFIFNHSEIILAQSS 373
Db 441 ELQKQBELRQOELVLEQEAWRLRRVVELQOGDSAQGQKEQOQEFELH-----LAVRE 495
QY 374 RNLHEV-----GRKSVKSGPVKNPHKRSH-----ELSFSDA 406
Db 496 RRELQMLNGLEAKQSELSELITREALESSHLSGELLRQEQTEVTAALARAQSI AEL 555
QY 407 PSNESLYR-----NETVAI-----SPPKKQQRVEDTTLPR 435
Db 556 SSSENTLKTVEADLRAAAVKLSALNEALALDKVGLNQQLLEENQSVCSMEAAEQAR 615
QY 436 SAAPAN-ASARSSSAHAW---PHAQSKFKLAQWMSLMKKPATPKVVRTSIQPSVKLYD 491
Db 616 NALQVDLAEAEKRREALMEKNTHLEAQLOKAEAGAEIQ--ADLRDIOBEKEIOKKUSE 673
QY 492 SELYMQTCINPDPPFAATTIDPF-----LASTWYLDE-----QAVDR 528
Db 674 SRHQE-----AATTQLEQLHQEAKQOEVLARAVQEKALVREKAALEVRLOAVER 725
QY 529 HQADFKKMINALVSI PADLDADL-----NNKIDVGLFNEVRNKLNVVAPTKESQSMY 582
Db 726 DRQDLAEQQLGLSSAKELLESSLFEAQOQNSVIEVTKGLEVQIQTVTQKEVIOGEVRC 785
QY 583 LTKYLETLRK-----AAVELFFSEQ--WRLPCSKVA--VYVYKQALAIRSDNLH 629
Db 786 L-KLELTERSQAQERDAAARQLAQAEQKGTALAEQQAHEKEVYQREKWEKERSWH 844
QY 630 LDWV-----MORTILELILCFNPLWRLGLEVVFGEKIQMQRNDIVGLSTFILNRLFR 683
Db 845 QQELAKALESLEKWE-----LEMRLKEOQTEMEAIQAO----- 880
QY 684 NKCEBQRYSKAYTLTEEVAETIKKHSLOKILFILPLDQAKO-----KRIVKHNPCLFVK 738
Db 881 ---EEERTQAESALCQMLETEK---ERVSLLETILQTKELADASQOELERLQRDMKVQ 933
QY 739 KSPHKETKDILRFSSELLANIGDITRELRLRGYVLQHRQTFLDEDFDYAFNNLAVDLRQ 798
Db 934 KLKEQETTGIQTLOEAQRELUKEAARQHRDDIALAQEBSL-----LQDK 980
QY 799 VRLTRVVEVI-----LLRDDLTRQLR-VPAISRLO-----IFNVK 833
Db 981 MDLQKQVEDLKSQVAQDDSORLVQEQVEKLERETQYNNRIQKELEREKASLTLSMEKE 1040
QY 834 LALGALGEANFOLGGDIAA--QDIVDGHRE-KTSLMLMOLI---YKFRSPKTHA-AATVL 886
Db 1041 QRLVLQEADSIIRQELSALRQDMQEAQGEQKELSAQELLRQEVKKEADFLAQAQL 1100
QY 887 QKWRRHMLHVVIQRRIRHKELMRHRANTVIOAIFRGHQMRYVKLFTX-RTQAAIL 945
Db 1101 EBLEASH-----ITEQQLRASLWAQAKAQL-----QLRLRSTESQLEALAAE 1144
QY 946 QKFTRRYLAQKOLYQSYHSIITIQWRRAQQLGRQHRQFVELREAAIFLQRIW----- 999
Db 1145 QQFGNQQAQAQALASLYSAL-----QQALGSVCESR-PELSGGDSAPSVWGLEPQ 1195
QY 1000 --RRLFAKKLLAAAEATRLORSQKQAAASYI---QMWRTYQLGRTORHEFLRQDL 1053
Db 1196 NGARSILFKRGPLLTALSA-----EAVASALHKLHDLWKTQ-----TRDV 1236
QY 1054 INFVQRMESKWSMLEQRK-----PFQOLKRAINIQQ---RWRAKLSMKCNADYIALR 1105
Db 1237 LRQDVQKLEFLRLTDEAKSQVHTLQDLQRLSQNQEESKWEKG--QNSLESELMELH 1294
QY 1106 SSVLKVQA-YRKATIQ---MRIDRNHYSLRNKVICLOQLR-AIMKMEQ----- 1151

```

```
Db 1295 ETWASLQSLRRABLRQWMAQGERELLQAAKENLTAQVEHLQAQAVVBARAQASAAAGILEE 1354
Qy 1152 ----RENYLRNRASILVQRYRMRQWMIQDR-----NAYLRTRKCIINVQR 1194
Db 1355 DLRTARSAALKDKNEV---ESERERAAQALQBGELKVAQKALQENLALITQ-----1403
Qy 1195 RWRATLQRRERKNYLHLQTTTKRIQIKFRAKRMKQARF--LQLKKVTLVVQKERRA 1252
Db 1404 ----TLAEREE-----EVELTGGIQ-ELEKQREMQAALELLSLDLKKNQEVDLQOBQ 1453
Qy 1253 LLQMRKQREYVHLREVTIKLQRFHQAQKSMRFMRKYRGTOAAV---SCLQMHWRNHL 1309
Db 1454 IQELEKCRSVLEHLPMVAQREQLTVQREIQIRELEKDRQTRQNVLEHQLLELEKKQMI 1513
Qy 1310 KRERNFSFLQRAAITIQ-----RRYRARLNMKOLKSYAQKQAAIT-----I 1354
Db 1514 -ESRQGVQDLKQGLVTLLECLALEENHHKWCQCKLILKEGQRTQVALTHLTIDL 1572
Qy 1355 QTRYRAKKAMQKV-----VL-----YQKQREAIKVQRR---YRGNLEMRK 1393
Db 1573 EERSQELQAQSSQIHDLESHSTVLARELQERDQEVKVSQREQIEELQKHEHLTQDLERRD 1632
Qy 1394 QIEVYQKQRAVILQKWRISIRMDLCKAGYRIRLS-----SLSIQ-----KW 1439
Db 1633 QELMLOQERIQVLEQDQ---TROTKILEEDLEQIKLSLRERGRELTQROLMOERABEG 1688
Qy 1440 RATVQARR---QREIFLSTIRKVRUQAFI-----RATILM 1472
Db 1689 KGPSKAQGSLEHMKLILRDKKEVBCQEHHELBELKQLEOQLOGLHKVKVGETSLLL 1748
Qy 1473 RQRRREFEMKERRAAVVIQRRFRARCAMLKARQDYQLIQSSVILVQKRFMRNSMKQARQE 1532
Db 1749 SQREQEI-----VVLQOQLQ-----EAREQELKEQS--LOSQLEAQALAQORDQE 1793
Qy 1533 FVQLRTIAVHLQKFRGRLMIEORNCPLLRCMPGFQARA-----RGFMARKRFOA 1595
Db 1794 LEALQ---QROQQAQGOEERVKK-----ADALOGALEQAHTLKERHGELODHKEQA 1843
Qy 1586 LMTPEMDLIRQKAAKVIQRYWRYGILRRKQKHGGLDIRKTAQLRQE-----1635
Db 1844 RLLEE--ELAVEGRVQALEVGLDLRAESREQEKALLAQOQCAQEAQEHVETRALQD 1901
Qy 1636 ----AKAVNSVRCKVQEA VR 1651
Db 1902 SWLQAQAVLKERDQLEALR 1921
```

## RESULT 10

```
US-10-205-823-419
; Sequence 419, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Subhang
; APPLICANT: Wansley, Angela M.
; APPLICANT: Ghatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
```

```
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-205-823-419
```

## Query Match

2.78; Score 260; DB 14; Length 1979;  
Best Local Similarity 18.08; Pred. No. 1.le-10;  
Matches 303; Conservative 287; Mismatches 624; Indels 472; Gaps 69;

```
Qy 291 LRKYSESMKDL-SLLSPQTKYAIQGSMPNLNEM-----KIRSTEQNRVYQEQIQIKA 342
Db 258 LSDYEERIELENLQGGSGSVIETDLSKIYEMQKTIQVLQIEKVSTK--KMQLEDKI 315
Qy 343 KDLNSSSSSRAS-----LAGQOEFL-FNHSEILAQSSRFNLHEVGKSVKSGPVKNPKHR 397
Db 316 KDINKKLSSAENDRDILRRREQQLNVKQIMBECENKLE---CSKLQPSAVKQSDTMT 372
Qy 398 SHELSFSDAPNSLSYR-----NETVAISPPKQR--VEDTTILPRGAAPANASARS 446
Db 373 EKERILAQASVEVFFLQALSDAENETMRLSLNODNSLAEDNLKMKRIVLEKEKS 432
Qy 447 SSAHAWPHAQSKFKLAQTMSLMKKPKATPRKVRDTSTQPSVKLYDSBYLMQTCINPPFA 506
Db 433 LLSQKEELQSLKLKIANNEYVIKSTAT---RDISLDS--ELHDLRLNLE-----A 478
Qy 507 ATTTIDPFLA--STMVLDQAVDRHQADFKKWL-----NALVSPADLADLIN-- 552
Db 479 KEQELNQSISEKETLIAEIBELDRQOEATKHMILINDQLSKQONEGDSIISKLKQDLND 538
Qy 553 -----NKDIDVGLFNEVNRKELAVAPTKEEQSMN--YLTQYRLE----- 589
Db 539 EKKRVHOLEDDKMDITKEL-DVOKEKLI-----QSEVALNDHLTKQLEKDVENLVQDLN 593
Qy 590 -----TLRKAAVELFFSEQMRLPCSKVAVYVKNQAL-RIRSDRNHLHLDVVMQRTILE 641
Db 594 KSQESNVSIQENLEL-----KEHIRQNEEELSRI RNELMQSLAQDSNSNEKDT 642
Qy 642 LLCNPLWRLGLHVEVFGKIQMSNREDIVGLSTFILNRLFRNKCEQR----- 690
Db 643 LLKEREABVR-NLKONLSELEQLNENLKVAFDVYKMEKLVACEDVRHQLEECLAGNN 701
Qy 691 -----YSKAVTLTE--BYAETIKKHS-----LQKIL 714
Db 702 QLSLEKNTIVETLKMEKGEIEAELCWAKKLLBEANKYEKTEBELSNARNINTSALQIEH 761
Qy 715 FLAPFLDAQKRIKVNHNPCLFVKKSPHKETKIDILLFSSSELLANIGDITRELRGLYVL 774
Db 762 EHLIKLNQKKDMETAEELKKNIEQMDTDHKTQDVL---SSSL-----EQKQLTQLI 810
Qy 775 QHRQTFLDFDYAPNNLAVDIRDGVRLTRVVEVI----- 808
Db 811 NKKEIFIEKLKRSKSLQBELDKYSQALRKNEILRQTIIEKDRSLGSKMKBENNHLQBELE 870
Qy 809 LLRDDLTQLRVPAPISRLQIFNVVKALGALGAENFOLGGDIA-AQDIVDGHREKTLSSL 867
Db 871 RLREEQSRTPADVPKTLDSVTELASVQSINTKEHLEBEIKKHQKIEDQONQSKWQLL 930
Qy 868 WQL-----IYKFRSPKPHAAATVLI----- 886
Db 931 QSLQEQKKEMDEFRYQHEQMNAHTQLFLEKDEBEIKSLQKTIQKTLQHEERQDICTDN 990
Qy 887 -----QKWWRRHLHVVIQRR-----IRHKELMRHRRAATVIOAVFRGHQMR 928
```

Db 991 SDFQETKVSQSLNENGSEKDLKSAETERLVKIGERELIKLEKNISLTKQIDQLS 1050  
 QY 929 K-----YVKLFKTERTOQAIILOKFTFRYLAQKQ----- 957  
 Db 1051 KDEVGKLTQIIQOKOLEIQAHLARISSTSHQDVVYLOQOQAYAMEREKVFVAVLNKTR 1110  
 QY 958 ----LVQS VHSIITITQWRRA-----QQLGROHOFVE-----LREAAFIQRTWRRL 1003  
 Db 1111 ENSHLKTEYHKMDVAAKEAALIKLOENKKLSTRFSSGQDMFRETIONLSRIREKD 1170  
 QY 1004 FAKKILAAETARLQRSQKQAAASYIQMWRTYOLGRIQRHEP---LRQDLIMFVQR 1060  
 Db 1171 IEIDAL-----SOKCQTLLAVLQTSSTGNEAGGVNSHQFEELLQERDKLQOVVK 1220  
 QY 1061 MRSKWSM-----LEORKEFQOLKRAANIQORWAKLSMRKCNADYIALRSS 1107  
 Db 1221 M-EBWKKQVMTTVONNQHESAOLELHQL-QAQLVDSDDNNSKLQV-----DYTGL----- 1270  
 QY 1108 VLKVOAYRK-----ATIQMRIDR---NHVYSIRKNVICQOURLRAIMKREORE 1153  
 Db 1271 ----IQSYENETKLNFGQELAQVHISIQOLCNTKDLKLDIISPQLSSASLITPQSA 1327  
 QY 1154 NYLRLNASILVQGYRMQOM-----IQDRNAVLRTRKCIINVQRWR-----ATLQ 1201  
 Db 1328 ECLRASKSEVISESSSELLQLEELRKSLOEKDATITLQ-----ENNRLSDSIAATSE 1382  
 QY 1202 MRERKNYHLQTTTKRIQIKFRKREMKQKQABFLO--LKKVTLVQKRRALLQMRKE 1259  
 Db 1383 LERKE-----HEQTDSEIKOL-----KERQDVLOKLLKEDLLIKAKSQQLSSNEN 1429  
 QY 1260 -----RQEYHLREVTIKLQRRPHAKSMRFRMAK-YRGTOAAVSCLO--MEW 1304  
 Db 1430 FTNKVNEELLQAVTNLKERLILLEMIDIGLKGNEKIVEYRGKETYQALQETNMKF 1489  
 QY 1305 RNHLRKRENSFLQLRQAATLQRRYRARLNMIKOLKSYAQLKQAAITIQTRYRAKAM 1364  
 Db 1490 -SMWLREKEFECH-SMKEKALAFELQKKE-----KEQKGTGELNQL-----LNAVXSM 1535  
 QY 1365 QKQVLYIQKQRAI-IKVORRVRGNLEMRKQIEVQKQOAVIRLOKWRSTRDMRL-QK 1422  
 Db 1536 QKTVVQOQERDQVMLAKQKMENTALQNEV---QRLRDKFRSNQOELERLNHLLSE 1592  
 QY 1423 AGVRRIRLSLSFORKWRATVOARRQETIFLSTIRKVRMLQAFIRATLLMRQORREFEMK 1482  
 Db 1593 DSYTREALAAERDEAKLRKVVTVLEKLVSSNAMENASHQSVQVESLOEQLN----- 1646  
 QY 1483 RRAAVVIQRRFRARCAMLKARQDYQLIQSSVILVQRFRANKSMKQARQEFVQLRTIAVH 1542  
 Db 1647 -----VWSKQDETALQLSV-----SQEQVKQYALSLANLQWVLEH 1682  
 QY 1543 LQKFRG-----KRLMIEORNCFOQLLRCSMPGFOARAGFMARKKPFQALMTPEMD 1593  
 Db 1683 FQOEBEKAMYSAELEKOKQOLIAEWKKAENLEGKVISLOEQLDEANALDSASRLT-EOLD 1741  
 QY 1594 LIRQKAAKAVIQRYMRGVIILRRKXOGHLL-DRKRIALQROEAKAVNSVRCKVQOAVRP 1652  
 Db 1742 -VKEQOIEB-----LKRQNELQOEMLDDVQKKIMSL-----ANSSEKVDKVL--- 1783  
 QY 1653 LRGRFI 1658  
 Db 1784 MRNLFI 1789

## RESULT 11

US-10-369-493-22734  
 ; Sequence 22734, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 22734  
 ; LENGTH: 1837  
 ; TYPE: PRT  
 ; ORGANISM: Schizosaccharomyces pombe  
 ; US-10-369-493-22734

Query Match 2.7%; Score 254.5; DB 15; Length 1837;

Best Local Similarity 18.4%; Pred. No. 2.7e-10;

Matches 338; Conservative 355; Mismatches 728; Indels 415; Gaps 85;

QY 203 DLEQDPA--TNKTFVQKHSSETINISLDT---LDCSRIDGQHPHTPLNKTTTI---VHAHTT 254  
 Db 159 DLKDALASTHQVLELQHTQOEKASLOIWTYEFELQKLTKQNSILENNNTWLSRELGQVND 218  
 QY 255 RALACIHBEESPPRTPTKSAIHD--LKRD--KLVGSPLRKYSES-----MKDLGL 303  
 Db 219 KLLS-LHQE--ASLEKSQLSQSLSDAVLEKDALQRKVSSLSQQFTESNLRYQNIVAELSE 275  
 QY 304 LSPQTKYALQSGMPNLNEMKIRSIQ-NRYIOE--QQIQIKAKDLNSSSS-----EASL 355  
 Db 276 MKQOYEF--QVSF---EKEISSQKQISELMWEKCEDCSLRLEKQNSGELEKLELAAQ 330  
 QY 356 AQOQFELFNHSEILAQ--SRFNLHEVGRKSVKSGPVNPKRKRSH-----ELSES 404  
 Db 331 SSFEQLSEHKEAEASLKSQINFLKEVSSLE-SQLKANERLRHVDEIEISDMSELKYS 389  
 QY 405 DAPSN-----ESLYRNETVAISPCKQKQVDETD-----TLPSAAPANAS 443  
 Db 390 NLLNNSMGFKGQSSVSDLYSERLY-----YKQYEQTCQERVERLQSY--NHVM 437  
 QY 444 ARSSSAHAWPHAQSKKFKLAQTMSLMKKPKATPRKVRDTSIQPSVKLYDSE---LYMOTC 499  
 Db 438 EKANLQHPLVKQOFRFAHQ-----REIVAMSEQYQKSLCCKAKSRYEQ-- 484  
 QY 500 INPDPPAAATTTIDPFLASTMYLDEQAVDRHQADFKNLALNALSIPA-DLDDADLNN---KI 555  
 Db 485 -----LETIFKDKCTENKHYEQETKOLARQOVLLHDLCEGILVGV 528  
 QY 556 DYVKLFNEVRNRELVPAPTKBEQSMNYLTKYRLETLRKAAVELFFSEQMRPLPSKVAVV 615  
 Db 529 DSRKKINSYVEKSL-----TEDETDTDQIISRLSVVFRNIR-ELQQQNQLLSA-----V 577  
 QY 616 NKQALRIRSDRNHLDD--VVMQRTILLELLCFNPL-----WLRGLGVVFGKIQM 664  
 Db 578 HELADRMEXDEKPDLDGAEIOBETLIKANETIDQTKMLEEVSDDQRYSLKERDFFRSIV 637  
 QY 665 QGNRDIVLGSTFILN-RLPRNKCEQORYSKATVLEEYAEETIKKHSLOKILFILFLPDQA 723  
 Db 638 QENKLLDAPATPSKLNNTNLIQTSYORSIRLEQLTNELES-----LKSIRN 688  
 QY 724 KQKRVKHNPCILFVKKSPHKETKIDLLRFSSELLANIGDITRELRLGVYLOHROFTLDE 783  
 Db 689 KEKFPKFEATSSIQLEKSNIQLOLTSITSRSLALEKLNDEKSL-----VLSERSK--DE 741  
 QY 784 FDYAFNNLAVDLRDGVRLTRVVEVILLRDDLTFRLEVPAISRQOR----- 828  
 Db 742 LDESYSKLSQELQ-----ASKKIEV-----QNVSSQSLSI-CNSQLEQSNHIVDNLSKNLLL 791  
 QY 829 -----IFNVKIALGALGEANFOLGDDIAAOD-----IVDGHREKTLISLLWLIYK 873  
 Db 792 TSVKDKLADLSNLSKSLSSLOQDNFHMKAQIESNQEYTAIVDSMNSRIELSNDL--R 849  
 QY 874 FRSPKFAAATVLOKWRHRHLVHVIQRRIRKHELMRRHRAATVIOAVFRGHQMRKYVKL 933  
 Db 850 VANSKLSSECSDDVRR-----LTLQNSFD---LREHQ-TLVLOLQSNITELKQDITL 896





QY 949 TRRYLAQKOLYOSYHSIITIORWRAQQLGROHROFVELREAAIFLORIWRRLFAKGL 1008  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
615 NRELDVKKKFEDIN-----QEFVKIKKNEILKR----- 644  
QY 1009 LAAETARLQSQKQAAASYIOMQWRTYOLGRIORHEFLRQDLIMFVQRRMRKWSML 1068  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
645 -----NLENTQ-NQIKAEYIS-----LAEHE-----AKWSSL 670  
QY 1069 EORKEFQOLKRAAINIQWRRAKLSMK-----CNADYLALRSSVLVQVAYRKATIOMRIDR 1125  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
671 SQ-----SMRKVQDSNAEILA-----NYRKG----- 691  
QY 1126 NHYSLRNKVICLOQLRAIMKREORENYLRNASILVQKRYMRQMIQDRNAYLRT 1185  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
692 -----QEEIVLHAEIKAKKELDTIQECIKVKYAPI-----VSF 726  
QY 1186 RKCINQRRWRATLQMRERNKYNLHLOTTTKRIQIKFRAKEMKORAEFLQLKKVTLV 1245  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
727 EBC-----ERKPKAT-----EKELKQQLSEQTKYSV-----SEEVKNKQENDKAKKEIFT 774  
QY 1246 VOKRRP-----ALLOMKEKEQYHLREVITKLORRFAOK-----SMRFMAKVRGTQA 1295  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
775 LOKOLRDKTVLIEKSHEMERALSRTDELNKLQKLSQKYTEVKNVKEKLVENAKOTSE 834  
QY 1296 AVSCICQHWNRNL-----LRKRERNSPLOLROAAITLORRYRARLNNMIKOLKSYAQ-L 1347  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
835 ILAVQNLQKHQVPLEQVEALKKSLNGTIENLKEHLKSMQRYEKQQTIVTKHQLLENQ 894  
QY 1348 KOAATITQRYRAKAMQKVLYOKOREBAIIKQVRRYRGNLEMRKQIEVYQORQAVIR 1407  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
895 KNSVPLAEHLQIKEAFEV-----GLIKA-----SLREKEESQNMKEEYSK 938  
QY 1408 LQKWRESITDMLCKAGYRRITLSLSIQKRWATVQVQRRQREIFLSITRKVRLMQAFIR 1467  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
939 LOS-----EVONTKOALKKQETREVPVDSLKYKAT---KSDLETQISSLINE----- 980  
QY 1468 ATLLMRQORREFEMKRAAVIQRFRACAMLKARODYOLLIQSSVILVQKFRANRSMK 1527  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
981 -----KLANLRKKEEVEEVLHAKKEISAKDKELHFSI-----EKEIK 1022  
QY 1528 QARQEVOLRTIAVHLOQKFRKRLMIQR-----NCFOLLRCSPGFGQ--ARARGF 1577  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1023 DOKERCDSLTTITELQRRIOESAKQIEAKONKITELANDVERLKAQALNGLSOLYTSGN 1082  
QY 1578 MARKRFOALMT-----PEMDLIRQKRAKVIQRYWGYLIRRRQKH--QGLDIR 1626  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1083 PTKRSQSLIDTLQHVKSLEQOLADADRQHEVIAI---YRTHLLSAAQGHMDEDVQBAL 1139  
QY 1627 KRIAQLRQ 1634  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1140 LQILQMRQ 1147

## RESULT 13

US-10-295-027-446

Sequence 446, Application US/10295027

Publication No. US2003023250A1

## GENERAL INFORMATION:

APPLICANT: Afar, Daniel

APPLICANT: Aziz, Natasha

APPLICANT: Ginsberg, Wendy M.

APPLICANT: Gish, Kurt C.

APPLICANT: Glynn, Richard

APPLICANT: Hevezi, Peter A.

APPLICANT: Mack, David H.

APPLICANT: Murray, Richard

APPLICANT: Watson, Susan R.

APPLICANT: Eos Biotechnology, Inc.

TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

FILE REFERENCE: 018501-012500US

CURRENT APPLICATION NUMBER: US/10/295,027

CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 446  
LENGTH: 1416  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-446

Query Match 2.6%; Score 250.5; DB 15; Length 1416;  
Best Local Similarity 18.1%; Pred. No. 3.7e-10;  
Matches 262; Conservative 243; Mismatches 522; Indels 421; Gaps 58;  
QY 277 IHDLRDKLVGSPURKYSESMDLSLLSPQTKYAIQSGMPNLNE--MKIRSIQNRYQ 334  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
296 IQDLFIENEDLKERLKIQQEQRIL-----LDKVNGLQLQNEEVMVADDESER--E 346  
QY 335 EQQIQIKAKDLNSSSSSEASLAGQOEFLEFNHSEIIAQSSRFN-----LHEVGRKSVKGS 388  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
347 KLSLLAAKEQHEESLRTIEALKNRFKYFSDHLSGSHFSNRKEDMLLKQGMVADS 406  
QY 389 PVKNP-----HKR-----RSHLSFSDAPSNESLYENETVALSPPKQORVEDTILPSAAP 439  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
407 QCTSGIIPAHQMSRMLRPLESL---PSQTSYSENEIL-----KKELEAMRTFCESAKQ 458  
QY 440 ANASARSSSAHAWPHAQSKFKLQATMSLMKPKPATPKRVDTSIQF-----SVKLYD 491  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
459 DRLKLQNELAH-----KVAECKALALECERVKEDSDEQIKQLEDALKDVQKRYE 508  
QY 492 SE---LYMQTCINPDPPFAATTITDFFLASTWYLDQAVDRHQADPKKWINALVSTPADLD 548  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
509 SEGKVKQMOT-----HFLALKEHLTSEAAAGN-----HRLTEELKDQL 546  
QY 549 ADLNNK-----IDVGLFNEVRNKELVVAPTKESQSMNLYTKYRLETLRKAARELVFPSEQ 603  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
547 KDLKVKYEGASAEVKGKRNQIKQNEMIVEEPFDEGK-----LIEENKRLQKLSMCEM 600  
QY 604 MRLPCSKVAVYVKNQALRIRSDRNHLHDVVVMQRTILLELLCFNPLWLRGLGVVFGKIQ 663  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
601 EREKKGKRVTEMGQAKLSA-----KSLAISIAEKEFN 634  
QY 664 MQSNRDIVGLSTFTILNRLFRNKCEQRYSKAYTTTEEYAEITIKKHSLOKILFLPFLDQA 723  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
635 MKSS-----LSNEVNE---KAKLVEMEREHEK--SLSEIRQLKRELENV 674  
QY 724 KOKRIVKHNPCLFVKKSPHKETKDILLRFSSEILANIGDITRELRRLGYVLOHRTFLDE 783  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
675 KAK-LAQH-----VKPEHEQVKSRLQKSGELGKKITELTKNQTLQKEIE--KVLYDN 726  
QY 784 --FDYAFNNLAVDLRDGVRLTRVVEVI-----LLRDDLTRQLRVPALSRLOIRFNVKIAL 836  
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
727 KLLKEQAHNLTIEMKNHYVPLKVSDEMKSHDAIIDDLNRKL-----LDVTQKYTEKKLEEM 782

QY	837	GALGEANFOLGGDIAAQDIV-----DGHEKXTLSLLMOLYIKFRSPFHPHAAATVLOK--WW	890
Db	783	EKLLENDLSKDVSRLETVFPPEXHEKEBIIALKSNIV-----ELKKQLSELKKKCGE	836
QY	891	RRWLHVVIORRIRHKELMRHRHAAVIOAVFGHQMRKVYKLFKTERTQAAI--ILOKF	948
Db	837	DQEKIHALTSENTNLKMM-----SNQYVPVKTHEEVKMTLNDTLAKT	879
QY	949	TRRYLAQKOLYQSVHSIIITIQRWRAQQLGRQHFVELREAAIFLQIRWRRLFAKKL	1008
Db	880	NRELLDVKKKFEDIN-----QEFVKIKDKNEILKR-----	909
QY	1009	LAAAEATLARQSRQKQAAASYIQMWNTYOLGRIORHEFLQRDLIMFVORRMRKWSML	1068
Db	910	-----NLENTO-NQIKAEYIS-----LABHE-----AKMSSL	935
QY	1069	EORKEFOOLKRAAINIOORWEAKLSMEK---CNADYIALSSVLKVQAYKATIQMIDR	1125
Db	936	SQ-----SMRKQVDSNAEILA-----NYRKG-----	956
QY	1126	NHYVSLRKNVICLQORLRAIMKMEQRENTYLRNLNASILVKQRYMRQMIODRNYLRT	1185
Db	957	-----QEEIVTLHAEIKAQCKELDTIQECIKVKYAPI-----VSP	991
QY	1186	KCIINVORWRPATLOMRERKNYVLIHQTOTTKRIQIKFRKREMKQORAEFLOIKVTLV	1245
Db	992	EEC-----ERKFKAT---EKELKDQLSQOTQKYSV---SEEEYKKNQENDKLKKEIFT	1039
QY	1246	VOKERR---ALLQMRKEROEVLHIREVTIKLORRPHAOK-----SMRTFRAKYRGTQA	1295
Db	1040	LOKDRLKTVLIEKSHENRALSRTDELNNKQLKLSQKTEVKVQVKEKLIVEENAKQTSE	1099
QY	1296	AVSCLOWHRNHL-----LRKERNSFQLRQAAITLQRRYRRLNMILKOLKSYAQ-L	1347
Db	1100	ILAVQNLLQXHVLEQVEALKKSLNGTIENTLKEELSMQRCYEKEQOQTVTKLHQLLENQ	1159
QY	1348	QQAAITQTRYRAKAMQOVVLQKORBALIKVORRYGNLEMRKQIEVYQQRQAVIR	1407
Db	1160	KNSVPLAEHLQIKEAPEKV-----GLIKA-----SLREKEEBSQNMKEEVSK	1203
QY	1408	LQKWRSIRDMRLCAGVYRIRLSSLSIQKRWATVQARRQREIFLSTIRKVRMLQAFIR	1467
Db	1204	LQS-----EVONTQKALKLETEVDVLSKYAT---KSDLETOISLINE-----	1245
QY	1468	ATLLMQORREFEMKRRAAVIORRFRARCAMLKARQDYOLIOSSVILVORKFRANFSMK	1527
Db	1246	-----KLANLNRKYEEVCEVBLHAKKKEISAKDBKELLHPSI-----EQEIK	1387
QY	1528	QARQEFVOLRTIAVHLQOKFRGKELMTQR-----NCFOLLRCMSPGFQ--ARARGF	1577
Db	1288	DQKERCDSKTITITTELORRQIESAKQTEAKDNKKTIELNDVRLKQALNGLSQLTYTSGN	1347
QY	1578	MARKRFQALMT-----PEMMDLIRKRAAKVIOQRYWRGYLIRRRQKH--QGLLDIR	1626
Db	1348	PTKQSQSLDITLQHVQKSLSQOOLADARQHQEVTAI---YTHLLSAAQGHMDEVDQEAL	1404
QY	1627	KRIAQLRQ	1634
Db	1405	LOIITMRO	1412

Qy	850	I A-----AQDVGHRREKTLISLLWOLI--YKFRSPK-----FHAAT 888
Dd	467	MAHIQVCMQTGALELHKLIVNSLSQOLFETQHSRMKEKRIRAAOQLGIEADMLDQEA 526
Qy	885	VLOKWRRRHWHVVQRIIRHKELMRHRAATVIQAVFRGHOMKYVKLFKTERQAALI 944
Dd	527	FMCIQEAKTWEEDLQRLEEFEE-GERERLORMADSASLSQQLEQVKKLTLLRQQOULEA 585
Qy	945	LQK-----FTFRYYLAQQLYQSVDHSIITIQWRWAQQLGRGHRQRFVELREAAILQRITW 999
Dd	586	LCOPHLDMKOLTTOTALOSREBGLDALOTHYDELOARCELOGEAASREDPTICL-LQ 643

RESULT 14  
US-10-408-765A-2125  
Sequence 2125, Application US/10408765A  
Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, Eoin D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Glenn, Gary M.

```

QY 1000 RRLFAKLLAAETAR-----LORSQKQQAAS-----YIOMQR 1035
Db 644 NEKILEAALQAAGKEELDRGARRLEEGTEETSETLEKLEBELAIGSGVEHLQOETA 703
QY 1036 TY--OLGRIORHEFLRDLIMFVORRMRSKWSMLEQRKEFQOLKRAAINIQRWRKLS 1093
Db 704 ALKQOMKI--KEQFLOQKQVWEYRRDATSKDQLISE-----LKAT 743
QY 1094 MRKNADYALRSLVLLVQVYKATIQMRIDNH--YYSLRKNVICLOORLAINMRREQ 1151
Db 744 RKRLDSELKELRQELMQVHG--EKRTAEAEKSLKREVAQVRQHMADLEGLQSAQKDERDE 802
QY 1152 RENILR-----LRNASILVOKYRWQMIQMDNAYLRPRKCIINVORWR--- 1197
Db 803 METHLSQSFQKEQWAVTEANEALKQIEELQ-----EARKAITQKQKQRLG 853
QY 1198 -----ATLOMRERKNVYLH-LQTTTKRIQIKPRAXR-----EMKKORA-----E 1235
Db 854 SLDTSQKEMTKKAYENAVGILSRRLQELAAKEADAEGLQRAQGGSSDSSIALHE 913
QY 1236 FLQAKKVTLVQKRRALLQMKERQEYLHL-----REVTIKLRFFHAQKMRMR 1287
Db 914 RIQALEAEALQAVSHSKTILE--KEIQEVIALTSQELSESREKVLSELEDEL--QESRGF-- 967
QY 1288 AKYRGTAAVSLQWNRNHLIRKBERNSFLQROAAITLORRYARLNMIKOLKSYAOL 1347
Db 968 -----RKTKRLEESN-----KKLALEHE--KGTGLGQ--SNAAL 1002
QY 1348 KOAATITOTRYAKKA-----MQQVVLQKOREALIKVORRYRGNLEMRQIEVYQ-- 1399
Db 1003 RHNSILFTALAKRADILVOLNLQVAVLQKKEE-----DRQKHLVQALQAS 1051
QY 1400 --KQAVIRLOKWRSDIMELCKAGYRIRLSLSIQKRWATVQARRQREIFLSTIR 1457
Db 1052 LEKEKEKNSLKEQVAAAK---VEAGHNRHFKAASLE-----LSEVK 1091
QY 1458 K-----VRLMOAFIRATILMRQORREFEMKRAAVIQRFRARCAMLKARQDYOLI 1509
Db 1092 KELOAKEHLVQKLQA--EADDLQITREGKHSQETAAQFOALAEA--RAQLQLOKQDLQJLS 1148
QY 1510 QSSV-----ILVORFRANRSMK-----QARQEFVQLRTIAVHLQOKFR--GK 1550
Db 1149 KQPVNGQENLKWVDQKEREIQLSKQQLDLTEQGRKEGLQQLQNVKSELEMAQE 1208
QY 1551 RLMIQRNCFQLLRSCMPGQARARGFMARKRQALMTPEMDLIRQKRAAKVIQRYWEG 1610
Db 1209 DLSMTQKDKFML-----QAKVSELKNNMKTLLQONQOLKLDLRRXXXXKGAER--RG 1258
QY 1611 YLIRRRQKHQ 1620
Db 1259 QLFQPCHAHQ 1268

```

RESULT 15

```

US-10-408-765A-1696
; Sequence 1696, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boon D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 1696
; LENGTH: 1893
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1696

Query Match
Best Local Similarity 2.5%; Score 238.5; DB 16; Length 1893;
; Mismatches 241; Indels 529; Gaps 79;
Matches 336; Conservative 241; Mismatches 592; Indels 529; Gaps 79;

QY 262 BEEGSPPTTKTAIHLDKDKILVSPFKYSEMKDLSILSPQTKVAIGSGMPNINE 321
Db 188 BESSQSPSGGLKSAHQWL-----AELREKLMQOATQOLGQOALLAGTPTKVEQ 241
QY 322 MKTISIEQNRYYEQIOIKAKDLNSSSSSSASLAGQOEFLE-----NHSEILAQSSRFN 376
Db 242 --ELRAIQOR--DOVYQTWAR-----KQELQAEQOQLFLRECGRLEILA----- 285
QY 377 LHEVGRK-SVKGSPVKNPHKR-RSHE-----LSPSDAPSNBSLYRNEVVAISPP----- 423
Db 286 AQEVSLSKTSALGSSVEVEVQLIRKHEVFLKVLTAQD--KKEAALRELRKTLRRPRVDR 343
QY 424 -----KKORVEDT-----TLPRSAAPANASARSSSAHAWPHAQSKKFKLAQTSIMMKP 472
Db 344 PILLQRRMRVKELAESRGHALHASLIMASFTQAAQAEADWIAQWAO-----LKEP 394
QY 473 ATPRKVRDTSIQPSVKLYDSELYMOTCINPDPPFAATTTIDPFLASTMYLDEQAVDRHQAD 532
Db 395 VPFGDLRD-KUKPLK-----HQA- 412
QY 533 FKKWLNALVSPADLDADLNKIDVGLFNEVKNELVVA---PYKEQSGMYLYTKYRLE 589
Db 413 -----PEAEVQAHEEV--MTSVAKGEALLAQSHPRAGEVSQ-----RLQ 450
QY 590 TIRKAAVELFSEQMRIPCCKVAVVYNKQALRIR-----SDRNHLHDVVMQRTILELLCF 645
Db 451 GURKEWEDL-----RQAMALRQGELEDNRNLEFQFQVDAEAWIOE 492
QY 646 NFLWLRLG--LEWVF-----GEKTMQSNREDIVGLSTFILNR----- 680
Db 493 KEVKNVGDIGDLQLEHCLQLRRRLREFRGNAGDVGDAICIRISDLSIQLAKNRDPEVK 552
QY 681 -----LFRNKCEQRYSKAYTLTEEYATIKKHSIQKILFLPFLD 721
Db 553 IICQRRSOLNWRWAFHGNLLRYQOQLEGALETHVLSRELDNVTKR--IQEKEALIQALD 610
QY 722 QAKQKRIKVNFCFLFKKSPHKEKTDILLRFSSELLANIGDITRELRRL-----GYVL 774
Db 611 CGKDLSEVQR-----LLRK--HEE-----LEREHPQAOQVESLEREVRGLCQSPAAHGL 660
QY 775 QHROTFLDEFDYAFNNLAVDLRDGVLRLTRVWEVILLRDDTLQRLRVPAISRQIRFNVKL 834
Db 661 RHROQVAVESWQLRSRAQKRREAL-----DALHQAKLQAM--LQELVSAQ 706
QY 835 ALGALGEANFOLGGDIAAQDIDVGHREKTLSL-LW-----QLI-----YKFRSPK 878
Db 707 RLRAQMDTSPAPRSPVEARRMLEEHOECKAELDSWTDISLARSTQQLLTAGHPFSSDI 766
QY 879 FHAAA-----TVLQKWRHHLH-----VVIQRIKHKLMRHRAATVIAQVFRG---- 924
Db 767 RQVLAGLEQLSLEGAWQEHQLOQOALELOFLSSVEKMERWLCSEKDSLASBGLWDP 826
QY 925 -----HOMRKY--VKLFKTERTOAAIILQKFTTIRYL-----AQKOLYOSYHS 964
Db 827 LAPMEPLLWKHKLMEWDLQVQACKISALEA-----TARGLHQGHPEAQAQALRCQAM 879
QY 965 IITIQRWRAQQLGRQ---HRQFVELREAAIFLQ-----RIWRRLFAKLLAAAE--- 1013
Db 880 LL-----RKEALFRQAGTGRHRLEBLRLQAFLOQDSQEVAAWLE---KNLVALEGL 930
QY 1014 -----TARLORSQKQAAASYIQWRTVQLGRIORHEFLRORDLIMFVORRMRSKWSML 1068
Db 931 DTAMLPAQLQKQONFOA-----ELDASHMQOQELORE---GQR-----LUQGGHPASEAIQ 978

```

Search completed: September 13, 2004, 17:56:15  
Job time : 206 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 14, 2004, 21:48:33 ; Search time 15801 Seconds  
(without alignments)  
5104.828 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMP 1861

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DBV=xlh  
-O=/cgn2\_1/USPTO\_spool\_h/US09914698/runat\_10092004\_150005\_28214/app\_query.fasta\_1.2055  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09914698 @CNC 1 1 1133 @runat\_10092004\_150005\_28214 -NCFU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	9514	100.0	5586	3	DMU95171	U95171 Drosophila
2	9386.5	98.7	57009	2	AC019775	AC019775 Drosophila
3	9386.5	98.7	215711	3	AC008205	AC008205 Drosophila
4	9386.5	98.7	228433	3	AE003749	AE003749 Drosophila
5	5267	55.4	3486	3	AY094825	AY094825 Drosophila
6	1374	14.4	9846	10	AF533752	AF533752 Mus muscu
7	1360	14.3	10434	9	AF509326	AF509326 Homo sapi
8	932.5	9.8	3760	9	BC034607	BC034607 Homo sapi
9	834	8.8	2453	6	AX876627	AX876627 Sequence
10	834	8.8	2453	6	BD156229	BD156229 Primer fo
11	834	8.8	2453	6	AK001380	AK001380 Homo sapi
12	755	7.9	4088	9	AK125107	AK125107 Homo sapi
13	753	7.9	235179	2	AC097078	AC097078 Rattus no
14	744.5	7.8	62965	9	AY365047	AY365047 Pan trogl
15	737	7.7	5760	9	HSM808955	EX648804 Homo sapi
16	735.5	7.7	2474	9	AY099893	AY099893 Homo sapi
17	735.5	7.7	2477	9	AY099891	AY099891 Homo sapi
18	721.5	7.6	155892	6	AL353809	AL353809 Human DNA
19	707.5	7.4	3887	6	AX876625	AX876625 Sequence
20	707.5	7.4	3887	6	BD156228	BD156228 Primer fo
21	707.5	7.4	3887	9	AK001379	AK001379 Homo sapi
22	688	7.2	2847	9	AY099890	AY099890 Homo sapi
23	678	7.1	2811	6	AX833832	AX833832 Sequence
24	678	7.1	2811	9	AK095892	AK095892 Homo sapi
25	616.5	6.5	1950	10	BC022903	BC022903 Mus muscu
26	443	4.7	2336	9	BC040439	BC040439 Homo sapi
27	437	4.6	2414	6	AX876687	AX876687 Sequence
28	437	4.6	2414	6	BD156262	BD156262 Primer fo
29	437	4.6	2414	9	AK001411	AK001411 Homo sapi
30	386	4.1	193262	3	AC007579	AC007579 Drosophila
31	386	4.1	260367	3	AE003808	AE003808 Drosophila
32	384	4.0	160827	3	AC099025	AC099025 Drosophila
33	359.5	3.8	68727	3	AC004516	AC004516 Drosophila
34	352	3.7	309506	10	AY363102	AY363102 Mus muscu
35	350.5	3.7	36134	3	CEC45G3	Z92780 Caenorhabdi
36	347.5	3.7	6925	5	GGDILUTE	X67251 G.gallus di
37	338	3.6	6599	5	GGP190MHR	Z11718 G.gallus nr
38	334.5	3.5	71023	2	AC004426	AC004426 Drosophila
39	324	3.4	9121	4	OCTRICA	Z19092 O.cuniculus
40	324	3.4	308108	3	AE003473	AE003473 Drosophila
41	323.5	3.4	9344	4	OATRICH	Z18361 O.aries tri
42	322	3.4	14800	6	AX334119	AX334119 Sequence
43	322	3.4	14800	6	HSU53204	U53204 Human plect
44	318.5	3.3	144056	3	AC091222	AC091222 Drosophila
45	318	3.3	5970	3	PFLSALG	X56203 P.falciparu

ALIGNMENTS

RESULT 1

```

DMU95171      5586 bp      mRNA      linear      INV 09-APR-1997
LOCUS         Drosophila melanogaster microtubule associated protein (asp) mRNA,
DEFINITION    complete cds.
ACCESSION     U95171.1 GI:1930121
VERSION       Drosophila melanogaster (fruit fly)
KEYWORDS      Drosophila melanogaster
SOURCE        Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
ORGANISM      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 5586)
AUTHORS      Saunders,R.D.C., Avides,M.C., Howard,T.I.A., Gonzalez,C. and
              Glover,D.M.G.
TITLE         The Drosophila gene abnormal spindle encodes a microtubule
              associated protein that associates with the polar regions of the
              mitotic spindle
JOURNAL       J. Cell Biol. (1997) In press
REFERENCE     2 (bases 1 to 5586)
AUTHORS      Saunders,R.D.C., Avides,M.C., Howard,T.I.A. and Glover,D.M.G.
TITLE         Direct Submission
JOURNAL       Submitted (26-MAR-1997) Anatomy & Physiology, University of Dundee,
              Medical Sciences Institute, Dundee DD1 4HN, UK
FEATURES     Location/Qualifiers
              1..5586
               /organism="Drosophila melanogaster"
               /mol_type="mRNA"
               /strain="Oregon R"
               /db_xref="taxon:7227"
               /chromosome="3"
               /map="96A-B"
               /map="96A-B"
              1..5586
               /gene="asp"
              1..5586
               /gene="asp"
               /function="required for the normal function of the mitotic
               spindle"
               /standard_name="abnormal spindle"
               /codon_start=1
               /product="microtubule associated protein"
               /protein_id="AAB51540.1"
               /db_xref="GI:1930122"
               /translation="MELVSPVLVEACKETLQIDNRNFRKEVMIILKSKNPQVKNP
               RKFPTVGTQLQKSTPGAGTKWKSIVSNAVQKKRMSAAAPPKQTRWRVTPSRPA
               WAHPQAPLVEKNVYKTPDEEPTIISPPSLKNSLSPMPPTGNLDDVIDNLRTPLT
               ETRGQQTIFPDNLAAWPTPLKGNVKSANDMRPRITPDDELDQPATNKTDFVXH
               SETINISLDTLDCSRIDQPHTPLNKTTIIVHATHRALACIHEEGEPSPPTPTKSA
               ITHLDKRDILVSPSPKYSWKDLSLSPOTKVAIOGSMPLNEMKIRSI EQNRYVO
               EQOIQIKAKDINSSSSEASLAGOQEFIFNHSIILAOSSRENLHEVGKRSVKGSPVKN
               PHKRSHELFSADPSNLSLYNETVALSPKQKQVETTLPRSNAPANASARSSAH
               AWPHAQSKFLKAPTSMLKKPATPKVRDTSIQPSVKLYDSELMFQICINPDPFAAT
               TIDPFLASTMYLDQAVDRQADPFKWLNALVSI PADLDADLNNKIDVGLKLFNEVRN
               KELVVAPTKEQSNMYLTKYRLTIRKAAVELFFSEQMRLLPCSVAVYVVKQALRIS
               DRNLHLDVMVORTLELLCFNPLWRLGLEVFGEKIQMQSNRDI VGLSTFILNRLF
               RNKCEORYSAYTLTEVATIKKHSIQKILFLPUDQAKOKRI VKNPCLFVKKS
               PHKETKDILLPSSHELLANIGITFEILRGVLVGHQOTFLDEFDYA FNNLAVDLRDG
               VRLTRVSVILLHLDLTQLRVPALSRQLRIFNVKLALGALCEANFQLGDIQAQDIV
               DGHEKTLISLLWQLYKPSFKFAAATVLQKWRHRLHVVIIQRIISHKELMRHRA
               RATVIOAFRGHMRKYVKLFKFTERTQAAIILQKFTRRYLAQKLYQSHYSITTORWW
               RAQOLGROHQRFPVLEAREATFLQRIWRRLFAKLLAAETARLQRSQKQAAASYI
               OMQWETYOLGRIFQHEFRDLIMFVORMRSKWSMLEORKEFOOLKRAAINTQORW
               RAKLSMRKNQIALRSRLKQVARKATLOMRIDRNHYYSLRKNVILCLOLRRAIM
               KUREORENLYLRNASILVQKRYNRQMIQDRNAYLTRKCLINQVREWATLQMR
               ERKNYLIHLITQRIQIPRAKREMKQRAEFLQKQKVTLLVVKRRALLQMRKE
               YLHLEVTIKLQRRFHAQSMFPAKRYGTQAAVSCLOMHRNHLRKRNSFLQJ
               ROAATLQRRYLRNLMIKQLSYAQLKAAITIQTRYRAKAKMQKQVLYQKQBEAI
               LKQVRYGNLBNMRKQIEVYQORQAVIRLQKWRISIRDMRLCKAGYRIRLSLSIQ
               PKWATVQARRQREIFLSTIRKVRIMQAFIRATILMQORREFEMKRAAVVIORRFR
               ARCAMLKQARQQLIQSSVILYORKFRANRSKQARQEFVOLRTTAVHLQKFKRKL
               MLEQRNCFQLLRKCFMFQARGFMAKRRFOALMTPPEMDLIRQKRAKVIQRYWRG
               YLIRRRQHQGLDTRKIQARQBEAKAVNSVRCKVQAVRPLRGFRFTASDALAVLSQ
               LDRLSRTPVHLMWCSFEMSTFCYGINAQINRSVEYDKQLIERCSRIILNLRNSTTY

```

```

NTPEGLVLTIAQMLLRWCDDSEIENFTLCTLIWFAHCCKKKKI IHDYMTNPEAIM
VRETKKLVARKKMKONARKPPDWTSGRYKSQINFTPCSLPSLEPFGIIRYSPYTF
ISSVYAFDITLCKLQIDMF"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 5586
Score: 9514.00 Matches: 1861
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-914-698-1 (1-1861) x DMU95171 (1-5586)
Qy 1 MetGluLeuValTrpSerProValLeuValAlaCysLysGluThrLeuGlnLeu 20
Db 1 ATGGAGCTAGTGTGGAGCCCGCTCTTGGAGTGGCCCTGCAAGAGAGAGCTGCACCTAATA 60
Qy 21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40
Db 61 GACACCCGCAACTTCCGAAAGGAGTGATGATCATCTCAAGTCCAAGAGCAACACGACCG 120
Qy 41 ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60
Db 121 GTCAAGAACCCGCGCAAAATTTCTTACTGTGCGCAAGACCCCTGCAGCTGAATCCCGACA 180
Qy 61 GlyAlaGlyLysThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMet 80
Db 181 GGAGCTGGCAAGACAATGAAGAAGCGTGTATTCCTGCTGTGCGCAAGAAAGAGCGCATG 240
Qy 81 SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro 100
Db 241 TCTGCAGCAGCAGCGCCTCCCTCCAGCAGACATGGCGAGTGACTGTCTTCCTTCGCTCCC 300
Qy 101 AlaAlaTrpAlaHisProProProGlnAlaProLeuValGluLysAsnValTyrLysThr 120
Db 301 GCTGATGGGCGACATCCACCTCCACAGCTCCTCTTGTTCGAGAAAGAAATGTATACAGACT 360
Qy 121 ProGlnGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer 140
Db 361 CCAAGAAGAGCCCGGTATATCATATCACACAGCTCTGCAATCTGCGATTACACACTCTCACC 420
Qy 141 ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThrProLeuThr 160
Db 421 CCTATGACCCAGGAACCTACTCGAGCTGATTGACAATCTGCGATTACACACTCTCACC 480
Qy 161 GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr 180
Db 481 GAAACCCGTGGCAAGGACAAAGCTACCACTTTTCCGGGCAATCTTGGCAGCCTGGCCACA 540
Qy 181 ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThr 200
Db 541 CCAACACTTAAAGGGAATGTAAATCATGTCTTAATGTATATGCGCCCGCTCGATCACT 600
Qy 201 ProAspLeuGluAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu 220
Db 601 CCAGATGATCTAGAAGATCAGCTGCCACAAACAAACGTTTCGATGTAAAGCATTCGAG 660
Qy 221 ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr 240
Db 661 ACTATCAATATTCGTTGACACCTTGACCTCCAGACTCGATCGATGGACACCCGATACG 720
Qy 241 ProLeuAsnLysThrThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle 260
Db 721 CCCCTTAATAGACACACCACTATGTGATGCCAGCACACAGAGCTCTGGCCCTGTATT 780
Qy 261 HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280
Db 781 CATGAGGAGGAGGAGGACCAAGTCCACTAGGACGCCCAAGAGCGCCATACACGACCTG 840
Qy 281 LysArgAspIleLysLeuValGlySerProLeuArgLysTyrSerGluSerMetLysAsp 300

```

841 AAGAGGACATTAAAGTTGGTGGCTTACCCTTACGAAAGTATTCAGAGTCCATGAAAGAT 900  
301 LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsn 320  
901 TTGTCACTTCTATCGCCACAACTAAGTAGTGGCAATTCAGAGGCTATGGCTTAATCTAAAT 960  
321 GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGluGlnIleGlnIle 340  
961 GAAATGAAATCCGCTCGATCGAACAGATAGATATTAACAGAGCAGCAGATCCAGATA 1020  
341 LysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu 360  
1021 AAGGCCAAAGACTGAATAGCTCTCTAGTAGCGAGGCTAGTTTGGCCGCCACGACGAG 1080  
361 PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluVal 380  
1081 TTTCATATTCACACAGTAGGATCCCTCGCTCAGTCAGTCGTTTAAATCTCCATGAAGTA 1140  
381 GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu 400  
1141 GGTCCGAAGTCGGTGAAGGAAGTCCGGTGAAGATCCCTCACAGCGCCGCTCTCATGAG 1200  
401 LeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThrValAlaIle 420  
1201 TTGAGTTTTTCGGATGCACCTAGCACGAATCATTTGACCGCAATGAAACTGTAGCCATT 1260  
421 SerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla 440  
1261 TCCCTCTCTAAAGCAACGGGTTGAGACACTACTCTGCCAGAAAGTCAGCGCCGCCA 1320  
441 AsnAlaSerAlaArgSerSerAlaHisAlaTyrProHisAlaGlnSerLysLysPhe 460  
1321 AATGCATCTGCAAGAGCAGTAGTGCACGCTGGCCACACGCCCAATCCAGAGATT 1380  
461 LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp 480  
1381 AAGCTAGCAACAAACCATGTCATCTGATGAAGAGCCGCCACACCCAGAAAGTCAAGGAC 1440  
481 ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle 500  
1441 ACTAGATTACGCTTCCGTGAAGCTCTATGACTGGAGCTGTATATGCAGACGTCATC 1500  
501 AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr 520  
1501 AACCCGATCCATTGTGCAGCAACTAGCAAAATGATCCATTCTGGCATCTACCATGAT 1560  
521 LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysIleTyrLeuAsnAlaLeu 540  
1561 TTGGATGAACAGGCTGTGATCGTCATCAAGCTGACTTTAAAGAGTGTGTTAAATGCCCTT 1620  
541 ValSerIleProAlaAspLeuAspAlaAspLeuAsnLysIleAspValGlyLysIleu 560  
1621 GTCTCATACCGCTGACCTGGACCGAGATTAAATAACAAATAGACGTTGGTAAAGCTG 1680  
561 PheAsnGluValArgAsnLysGluLeuValValAlaProThrLysGluGlnSerMet 580  
1681 TTAAACGAGGTGCGCAACAAAGAGCTCGTGGTGGCTCCCCCAAGAGGAGCAGTCTATG 1740  
581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600  
1741 AACTACCTAAACGAAATACCGCTCGAGAGCTTCGTAAGCGGCTGTGGAGCTCTTCTTC 1800  
601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
1801 AGTAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATATGTCAACAAAGCAAGCTCTG 1860  
621 ArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
1861 CGCATCCGTAGGATCGTAATCTTCACCTAGACGTAGTATATGCAACGACCACTACTGAG 1920  
641 LeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValPheGlyGlu 660  
1921 CTGCTGCTTTGCTTCAATCCCTTTGGTGGCTTGGACTTGGAAAGTGTCTTTGGCGAG 1980

661 LysIleGlnMetClnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
1981 AAGATCCAGATGAGTCTAATCGAGACATTTGTAGGCTCAGACCTTTATCTCAATCGC 2040  
681 LeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
2041 TTGTTCCGAAATAAGTGTGAGAGCAGAGTACAGCAAGCATAACACTCACCGAAGAG 2100  
701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720  
2101 TAGCGGAGACCATTAAGAAGCACTCATTTGCAGAAATCTCTTTTGTGCTCTTCTC 2160  
721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
2161 GATCAAGCTTAAGCAGAACGCGCATCGTCAAGCAATCCCTGTTGTTGTTAAAAAGTCG 2220  
741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIle 760  
2221 CCACATTAAGAGACCAAGGATATTCTGCTGCGCTTCTCTGCGAGCTGCTCGCAACAT 2280  
761 GlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
2281 GGTGATATTACGCGGAACTTCGTGCGCTGGCTACGTTCTACAGCACCGCCAAACATTT 2340  
781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
2341 TTGGAGAGTTTCAATATGCTTCAACAACTTGGCTGTGGACCTTAAGAGATGCGTGAGA 2400  
801 LeuThrArgValValGluValIleLeuLeuArgAspLeuThrArgGlnLeuArgVal 820  
2401 CTAAACCGAGTCTGGAGGTAAATTTACTACCGATGATCTAACCCGCCAGTTTAAGGGTG 2460  
821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
2461 CCAGCCATCTCTCGCTTCCAGCGGATCTCAATGTAAGCTAGCTCTGCGGCACATTGT 2520  
841 GluAlaAsnPheGlnLeuGlyClyAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
2521 GAAGCCAACTTCCAGCTAGCGCGGACATCGCGCCCAAGACATCGTTGACGACATCGT 2580  
861 GluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
2581 GAGAAGACGCTTTCCTGCTCTGGCAACTATTATCAAAATCCGCTCGCCCAAGTTTCA 2640  
881 AlaAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGln 900  
2641 GCGCGGCCACCGTCTCCAGAAATGGTGGCGCGCTCACTGGCTGCACGCTTGTATCCAG 2700  
901 ArgArgIleArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
2701 CGTCGATTCGCCCAAAAGAGCTTATGCTCGCCACCGCGCGCTACTGTCTATTCAGGCC 2760  
921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
2761 GTGTTCCGTGGCCACCATGAGAAAGTACGTGAAGTTGTTTAAAGCGAAGCAGCTCAG 2820  
941 AlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960  
2821 GCCGCAATAATCTGCAAAAGTTACCCGCTCGGTATTTGGCCCAAGACAGCTATATCAG 2880  
961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
2881 AGCTATCACGATATTATCACCATCCAGCGCTGGTGGCGAGGCCCAACAATCGGAGGCGAG 2940  
981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArg 1000  
2941 CACGCCACGCGTTGTGAGCTCGAGAGCTCGGATCTGCTTTTTCAGCGGATCTGGCGG 3000  
1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
3001 CGAGCACTCTTTGCCAAAAAACTATTGGCGCGCGGGGAAACAGCACGACTTCAGCGATCG 3060

QY 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGly 1040  
DB 3061 CAAAAACAACAGGAGCTGCTAGTTATTAATTCAAATGCAATGGCAACGTTATCAGCTGGGC 3120  
QY 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
DB 3121 AGAATTACAGCACACAGAGTTCTCGCCAGAGGAGCCTCATCATGTTGTTTTCAGCGAGG 3180  
QY 1061 MetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
DB 3181 ATGCGAAGCAAGTGGAGTATGCTGGAGCAGCGAAGAGATTCCAGCAACTAAAGCGGTGCA 3240  
QY 1081 AlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
DB 3241 GCTATTAATATCCAAACCTGGCGAGCGAGCTTTCAATGAGAAAGTGCACAGCTGAT 3300  
QY 1101 TyrLeuAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
DB 3301 TATTGGCACTTCGTTCCAGCGTTCTTAAAGTTCCAGGCTTACAGAAAGCCACAATCCAG 3360  
QY 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
DB 3361 ATGAGATATGATCGTAACTACTATTCCTCGCAAAAAATGTTATCTGCTGCAACAG 3420  
QY 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
DB 3421 CGACTGAGGGCCATCATGAAAATCGCGGACAGAGGGGAAAATATATCTGAGGCTGCGAAT 3480  
QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
DB 3481 GCTTCAATACTAGTTCAAAACCGTACCGCATCGTCAACAAATGATCCAGATAGAAAT 3540  
QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
DB 3541 GCATATTATAACCCGCCAATATGATCATCAATGTCCAGAGCGCTGGAGAGCCACTCTG 3600  
QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
DB 3601 CAGATGGCTCGAGAAAGGAGAACTACCTTCATCTCCAAACGACAAACGAATTCAA 3660  
QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
DB 3661 ATCAAGTTCCGTGCCAAGCGCGAAATGAAAAGCAAGAGCGAGTTCTTTCAGCTGAAA 3720  
QY 1241 LysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
DB 3721 AAGTTACCTTGTAGTCCAGAAACGTGCGCGGGCTTTGCTGCAGATGCGAAAGAGCGC 3780  
QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
DB 3781 CAGAGTACCTTACATCTCGCGAGGTTACCATAAAGCTGCAGCGCAGATTCCATGCTCAA 3840  
QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
DB 3841 AAATCAATGCGGTTTCATCGAGCAAGTACCGCGGCACCCAGGCTGCTGTGAGCTGCGCTG 3900  
QY 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320  
DB 3901 CAGATGCATTTGGCGTAAATCATCTCTTAGGAAACGGGAGAGAACAGTTTCTTACAACTG 3960  
QY 1321 ArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGln 1340  
DB 3961 CGTCAAGCAGCAATTAACACTGCAGCGAGCGGTACCGAGCTCGTCTGAATATGATCAAGCAG 4020  
QY 1341 LeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAla 1360  
DB 4021 TTGAAGATTAAGCTGCGCCAGCTGAAAACAGGCAGCTATTACATTCAAACCCGATATAGACC 4080  
QY 1361 LysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleLys 1380  
DB 4081 AAGAAGGCATATGCAAAAGCAGGTGCTTTGTATCCAAAAGCAAGGAGCACTATTCAAA 4140  
QY 1381 ValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLys 1400

DB 4141 GTCAACAGCAGATACCGCGCAATCTGGAGATGAGGAAGCAGATTGAAGTCTACCAAAA 4200  
QY 1401 GlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeu 1420  
DB 4201 CAGCGCAGGAGTGCATCCGCTTGCAGAAATGTGGCGCAGTATACCGCAATCGCGGCTG 4260  
QY 1421 CysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArg 1440  
DB 4261 TGCAAGCGGCTACCGCAAGGATTCGACTCAGTTTCATTGAGCATTCACGCAAGTGGCGG 4320  
QY 1441 AlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArg 1460  
DB 4321 GCCACAGTGCAGGCTGCTGCCCAACGGGAGATCTTCTTAAGCACCATCCGCAAAAGTCGA 4380  
QY 1461 LeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGlu 1480  
DB 4381 CTTATGCGAGCATTTATCAGAGCAACTTTACTGATGCTCAGCAGCGCAGGAAATTCGAG 4440  
QY 1481 MetLysArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeu 1500  
DB 4441 ATGAAGCGAAGGCTGCGGTAGTGATCCAGCGCGGTTTCGTGCTCGCTGTGCTATGCTA 4500  
QY 1501 LysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPhe 1520  
DB 4501 AAGCGCAGACAGGATTTACCAATTAATCAATCTCTGTGATCTCTGGTGGCAGCGCAAAATC 4560  
QY 1521 ArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAla 1540  
DB 4561 CGTCTTAATCGCAGATGAAAGCAGGACCGCAGGAATTTGTCAGCTGCGTACTATTGCA 4620  
QY 1541 ValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPhe 1560  
DB 4621 GTCCATTGTCAACAAAAGTTCCGTGGCAAGCGTCTAATGATGAGCAGCGTAATGTGTTTC 4680  
QY 1561 GlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArg 1580  
DB 4681 CAATGCTCGCTGTTCCATGCGCGGCTTCAGGACAGTGCCTGGCGCTTTATGGCTCGC 4740  
QY 1581 LysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAla 1600  
DB 4741 AAAGATTCAGGCGCTGATGACACCCAGATGATGACCTCATCCGCCAGAGCGCGCC 4800  
QY 1601 AlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGln 1620  
DB 4801 GCCAAGGTTATACAGCTTACTGGCGGGCTTCTAATCCGACAGCTCAGAGCACCCAG 4860  
QY 1621 GlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsn 1640  
DB 4861 GGCTCTTGGATATCCCGAAGCGTATTGCTCAGCTGCGGCAAGAGGCAAGCGGTAAAC 4920  
QY 1641 SerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSer 1660  
DB 4921 TCTGTGCGTGCAAAGTCCAGAGCGGTGCTGTTCTTCGCGAGCGCTTATCGCATCA 4980  
QY 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu 1680  
DB 4981 GATCGTTAGCAGTGTAAAGTCAATTGATCGTCTTTCGCGCATGTGCTCCACACTGCTC 5040  
QY 1681 MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg 1700  
DB 5041 ATGTGGTGTTCGGAGTTTCATGCTCCAGCTTTTGTGATGCGCATCATGGCTCAGGCCATTGCA 5100  
QY 1701 SerGluValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArg 1720  
DB 5101 TCAGAGTGGATAGACGCTTATAGCGCTGACGCGGATCATCTTAATTTGSCCGC 5160  
QY 1721 TyrAsnSerThrThrValAsnThrPheGlnGluGlyLeuValThrIleAlaGlnMet 1740  
DB 5161 TACAATAGCACACCGGTGAACACGCTTCAGAGAGGCGGTTTGGTCACCATTTGCCAGATG 5220  
QY 1741 LeuLeuArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrp 1760



```

Db      5221 TTATTGCGTGGTGGCGAAGAGCAGTCAGATATTAACACCTTTGTGCAACCCCTCATTTGG 5280
QY      1761 ValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGlu 1780
Db      5281 GTATTGCCACCTGTCTCTAAAGAGCGAAAGATCATTCACGACTATATGACCAACCCAGAG 5340
QY      1781 AlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGln 1800
Db      5341 GCCATTATACATGTGGCGAAGCTAAGAAAGCTGTGGCCCGCAAGGAAAGATGACAGCAA 5400
QY      1801 AsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPhe 1820
Db      5401 AATGCAGCGAAGCGCGCGCAATGACAAAGTGGACGGTATTAAGACCCAAAAGATAAACTTC 5460
QY      1821 ThrProCysSerLeuProSerLeuGluProAspPheGlyIleIleArgTyrSerProTyr 1840
Db      5461 ACCCGGTGTTCCCTGCCAGCTTGGAGCCGAGCTTCGGAATCATTCGCTACATGCCCTAC 5520
QY      1841 ThrPheIleSerSerValTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMet 1860
Db      5521 ACGTATTATCTCGTTCGGTTACGGCTTCGATACGATTTTGTGCAAGCTGCAGATCGACATG 5580
QY      1861 Phe 1861
Db      5581 TTT 5583

RESULT 2
AC019775
LOCUS   AC019775
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC019775
VERSION   AC019775.1 GI:6665122
KEYWORDS HTG; HTGS PHASE2.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
          Adams,M. and Venter,J.C.
          Direct Submission
          Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
          This sequence was identified as CDW:10210906 by the submitter.
          For more information on this record e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.

FEATURES
             source
             location/Qualifiers
               1..57009
               /organism="Drosophila melanogaster"
               /mol_type="genomic DNA"
               /db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.:      0
Score:          9386.50
Percent Similarity: 96.32%
Best Local Similarity: 96.11%
Query Match:    98.66%
Indels:         69
Gaps:           2

US-09-914-698-1 (1-1861) x AC019775 (1-57009)

QY      1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrIleuGlnLeulle 20
Db      21900 ATGGAGCTAGTGTGGAGCCCGCTCTTGGGTGGCGCTGCAAGAGACGCTGCAGCTATA 21959
QY      21 AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro 40
Db      21960 GACAACCGCACTTCGAAAGAGGTGATGATCATCTACTCAAGTCCAAGACCAACCGCG 22019

```

```

QY      41 Val----- 41
Db      22020 GTCAAGTCTAGTCAAAAAATGCATTTGCCACCCCTTTTACTTACATACCCCTCTTTTGCAT 22079
QY      42 -----LysAsnProArgLysPhePheProThrValGlyLysThrLe 54
Db      22080 GTATCTCTTACCTTTCTACCTGAAGAACCCCGCGCAATTTCTTACTGTCCGCAAGACCCCT 22139
QY      54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValValSerAlaAlaVa 74
Db      22140 GCAGCTGAAATCCGCCACAGAGCTGGCAAGACAATGAAAGCGTGTATCCCGCTGTGT 22199
QY      74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94
Db      22200 GCAGCAAAAGAAAGCGGATGTCGACGAGCGCGCTCCCTCCCAAGCAGACATGGCGAGT 22259
QY      94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValG1 114
Db      22260 GACTGCTCCTTCGCTCCCGCTGCATGGGCACATCCACCTCCACAGCTCCTCTTGTGCA 22319
QY      114 uLysAsnValTyrLysThrProGlnGluProValTyrIleSerProGlnProArgSe 134
Db      22320 GAAGAATGTATACAGACTCCACAGAGAGCGGTATACATATCACCAGAGCTTCGCAG 22379
QY      134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154
Db      22380 TCTTAAGGAAATCTAAGCCCTATGACGCCAGGAAACCTACTCGAGCTGATTGACAACTCT 22439
QY      154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174
Db      22440 GCGATTCACTCTCACCGAAACCCGTGGCAAGACCAAGCTACCATTTTCCCGGACAA 22499
QY      174 nLeuAlaAlaTrpProThrProThrIleuLysGlyAsnValLysSerCysAlaAsnAspMe 194
Db      22500 TCTGGCAGCTGGCCACACCAACACTTAAGAGGATGTAAATCATGTGCTAATGATAT 22559
QY      194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214
Db      22560 GCGGCGCGTCTGAATCACTCCAGATGATCTAGAAGATCAGCCTGCCACAAACAAACGTT 22619
QY      214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArg1 234
Db      22620 CGATGTAAAGCAATTCGAGACTATCAATATTTCTGTTGGACACTTGGACTGCTCCAGGAT 22679
QY      234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisTh 254
Db      22680 CGATGGACAACCGCATACGCCCTAAATAAGACAACAACCATTTGTGCAATGCCACGCACAC 22739
QY      254 rArgAlaLeuAlaCysIleHisGluGluGlyProSerProProArgThrProThrIly 274
Db      22740 CAGAGCTCTGGCTGTATTTCATGAGAGAGAGGAGCAAGTCCACCTAGACGCCACGAA 22799
QY      274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294
Db      22800 GAGCGCCATACACGACTGAAGAGGGAATTAAGTTGGTGGTTCCCTTACCTACGAAAGTA 22859
QY      294 rSerGlnSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnG1 314
Db      22860 TTCCGAGTCCATGAAGATTTGTCACTTCTATCGCCACAACTAAGTATGCCATTCAAGG 22919
QY      314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTy-G1 334
Db      22920 GTCTATGCTTAATCTAAATGAAATCGCTCGATCGAACAAGATAGATATTATCA 22979
QY      334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaSe 354
Db      22980 GGAGCAGAGATCCAGATAAGGCCCAAGACTTGATAGTCTCTTAGTAGGAGGCTAG 23039
QY      354 rLeuAlaGlyGlnGlnGluPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374
Db      23040 TTTGGCCGGCAGCAGGAGTTTCTATTCAACACAGTGAAGTCTCTCGCTCAGTCCAGTCG 23099

```

QY	374	gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi	394	Db	24180	TTTTTTTGTAAAGATCGCCACATAAAGAGACAAGATATTCTGTGCGCTTCTCGTC	24239
Db	23100	TTTTTAATCTCATGAAGTAGTGGAAAGTGGTGAAGGAGATCCGGTGAAGATCTCTCA	23159	QY	754	rGluLeuLeuAlaasnileGlyAspIleThrArgGluLeuArgLeuGlyTyrValle	774
QY	394	sLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrAr	414	Db	24240	GGAGCTGCTCGCAACATTGGTATATTACGGGAACTTCGTGCGCTGGGTACGTCT	24299
Db	23160	CAAGCGCGCTCTCATGAGTTGAGTTTTCGGATGACCTAGCAACGAATCATTTGTACCG	23219	QY	774	uGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValas	794
QY	414	gAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr	434	Db	24300	ACAGCACCGCAACATTTTGGACGAGTTTCGATATATGCTTCAACAACTTGGCTGGA	24359
Db	23220	CAATGAACCTGTAGCCATTTCCCTCTCTAAAGAACCGGTGAGACACTACTCTGCC	23279	QY	794	pLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAspLe	814
QY	434	oArgSerAlaAlaAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTyrProHi	454	Db	24360	CCTAAGAGATGGGTGAGACTAACCCGAGTCATGAGGTATATTTACTACGGGATGAT	24419
Db	23280	CAGAAGTGCAAGCGCGGCAAAATGATCTGCAGAGAGAGTAGTCCCAAGCTGGCCACA	23339	QY	814	uThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLe	834
QY	454	sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh	474	Db	24420	AACCCGCGAGTTAAGGTGCAGCCATCTCTCGCTTCAGGGATCTTCAATGTAAAGCT	24479
Db	23340	CGCCCAATCCAGAAGTTTAAAGCTAGCACAAACCATGTCTGATGAAGAGCGCGCAC	23399	QY	834	uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAs	854
QY	474	rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLe	494	Db	24480	AGCTCTGGCGCACTTGGTGAAGCAACTTCCAGCTAGGCGCGACATCGCGGCCAAGA	24539
Db	23400	ACCAAGAAAGTCAAGGACACTAGCATTCAGCCTTCGGTCAAGCTCTATGACTCGAGCT	23459	QY	854	pIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPh	874
QY	494	uTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrIleAspProPh	514	Db	24540	CATCGTTGACGGACATCGTGAGAAAGCGTTTCCCTGCTCTGGCACTTATTTTCAAAAT	24599
Db	23460	GTATATGACAGCTGCATCAACCCGGATCCATTTGCAGCACTACGACAAATGATCCATT	23519	QY	874	eArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTr	894
QY	514	eLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy	534	Db	24600	CCGCTCGCCCAAGTTTCATGGCGGGCCACGGTCTCCAGAAATGGTGGCGCGCTCACTG	24659
Db	23520	TCGGCATCTACCATGATTTGGATGAACAGCTGTGGATCGTCATCAAGCTGATTTAA	23579	QY	894	pLeuHisValValIleGlnArgArgIleArgHisLysGluLeuMetArgArgHisArgAl	914
QY	534	sLysTrpLeuAsnAlaIleValSerIleProAlaAspLeuAspAlaAspLeuAsnLys	554	Db	24660	GCTGCAGCTTGTATCCAGCGTCGCACTTCGCCACAAAGAGCTTATGCTGCCACCGGCG	24719
Db	23580	AAAGTGTAAATGCCCTTGCTCATACCCGCTGACTGCGACGAGATTTTAAATACAA	23639	QY	914	aAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPh	934
QY	554	sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValAlaProTh	574	Db	24720	CGCTACTGTCTATTGAGCGCGTTCGCGGGCCACAGATGAGAAAGTACGTGAGTTGTT	24779
Db	23640	AATAGAGCTTGGTAAAGCTGTTTAAAGAGGTGGCAACAAAGAGCTCGTGGTGGCTCCAC	23699	QY	934	eLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgTyrLeuAl	954
QY	574	rLysGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAl	594	Db	24780	TAAAGCGAAGCGACTCAGCGCGCAATAATTTCTGCAAAAGTTTCCCGCTGGTATTGGC	24839
Db	23700	CAAGAGAGAGCAGTCTATGAATCACTTAACGAAATACCGCTGGAGAGCGCTTCGTAAGGC	23759	QY	954	aglnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTyrTrpArgAl	974
QY	594	aAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTy	614	Db	24840	CCAGAAGCAGCTATATCAGAGCTATCAGATTTATCACCATCCAGCGCTCGTGGCGAGC	24899
Db	23760	GGCTGTGAGCTCTTCTTCAGTGAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATA	23819	QY	974	aglnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh	994
QY	614	rValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMe	634	Db	24900	CCAAACACTGGGAAGGCGACCGCGCTTGTGGAGCTCCGAGAGGCTGCGATCTT	24959
Db	23820	TGTCACACAGCAAGCTCTCGCATCCGTAGCGATCGTAATCTTCACTTAGACGTAGTTAT	23879	QY	994	eLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluTh	1014
QY	634	tGlnArgThrIleLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLe	654	Db	24960	TCCTCAGCGGATCTGGCGCGACGACTCTTTGCCAAAAAACTATTGGCGGGCGGAAAC	25019
Db	23880	GCAAGCACCACCTACTGAGCTGCTGCTTGTCTCAATCCCTTTGGCTGGCGCTTGAGCT	23939	QY	1014	rAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaIleAserTyrIleGlnMetGlnTr	1034
QY	654	uGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe	674	Db	25020	AGCCAGACTTCAGCGCATCGCAAAAACAACAGCAGCTCTAGTTATATTCAAATCAAATG	25079
Db	23940	GGAGTGTGCTTTGGCGAGAGATCCAGATGCACTCTAATCCAGACATTTAGGCTCAG	23999	QY	1034	pArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuI	1054
QY	674	rThrPheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAl	694	Db	25080	CGGAAGCTATCAGCTGGCGAGATTCAGCGACAGCAGTTCCTGCGCCAGAGGGACCTCAT	25139
Db	24000	CACCTTTATCCTCAATCGCTTGTCCGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC	24059	QY	1054	eMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPh	1074
QY	694	aTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLe	714	Db	25140	CATGTTTGTTCAGCGCAGGATGCGAAGCAAGTGGAGTATGCTGGAGCAGCGAAGAGATT	25199
Db	24060	ATPACACACTCACGGAAGATACCGCGAGACCAATTAAGAGACACTCATTCAGAAAATCCT	24119	QY	1074	eGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMe	1094
QY	714	uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCy	734	Db	25200	CCAGCAACTAAAGCGTGCAGCTATAAATATCAACACCGCTCGCAGCGAGCTTTCAAT	25259
Db	24120	CTTTTGTGCTTTTCTCGATCAGCTAAGCAGAGCGCATCGTCAACACAAATCCCTG	24179	QY	1094	tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAlaTy	1114
QY	734	sLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSe	754				

Db 25260 GAGAAAGTCACAGCGCTGATTATTTGGTTCCTGCTCCAGCGCTTTCTTAAAGTTCAGGCTTA 25319  
QY 1114 xArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
Db 25320 CAGAAAGCCACCAATCCAGATGAGATAGATCGTAATCACTACTATTCTCCGCGCAAAAA 25379  
QY 1134 nValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAs 1154  
Db 25380 TGTATTCTGCTGCAACAGCGACTGAGGGCCATCATGAAATGCGCAACAGAGGAAAA 25439  
QY 1154 nTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnG 1174  
Db 25440 TTATCTCAGGCTCGCAATCTCTCAATCTAGTTTCAAAACGCTACCGCATGGTCAACA 25499  
QY 1174 nMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnAr 1194  
Db 25500 AATGATCCAGATAGAAATGCAATTTAAAGAACCCGCAATGTATCATCAATGTCAGAG 25559  
QY 1194 gArgTTrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnTh 1214  
Db 25560 GCGCTGGAGAGCCACTCTGCAGATGGCTCGAGAAAGGAGAACTACTCTTCATCTCCAAAC 25619  
QY 1214 rThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl 1234  
Db 25620 GACAAACCAACGAAATTCAAATCAAGTTCCGTGCGCAAGCGCAAAATGAAAAAGCAAGAGC 25679  
QY 1234 aGluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLe 1254  
Db 25680 CGAGTTTCTTCACTGAAAGAGGTACCTTGTAGTCCAGAAACGTCGGCGGGCTTTGCT 25739  
QY 1254 uGlnMetArgLysGluArgGlnGlnLysLeuHisLeuArgGluValThrIleLysLeuGl 1274  
Db 25740 CGAGATGCGAAAGAGCGCCAGGAGTACCTACATCTGCGCGAGTTACCAATAAGGTGCA 25799  
QY 1274 nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl 1294  
Db 25800 GCGCAGATTCCATGCTCAAAATCAATGCGGTTCATGCGAGCCAAAGTACCGCGGCCACCCA 25859  
QY 1294 nAlaAlaValSerCysLeuGlnMetHisTTrpArgAsnHisLeuLeuArgLysArgGluAr 1314  
Db 25860 GCGTGTGTGAGTGGCTGCAGATGCAATTTGGGTAACTATCATCTCTTAGGAAACGGGAGAG 25919  
QY 1314 gAsnSerPheLeuGlnLeuArgGlnAlaIleThrLeuGlnArgArgTyrArgAlaAr 1334  
Db 25920 GAACAGTTCTTACAACCTGCGTCAAGCAGCAATACACTGCAGCGACGGTACCGAGCTCG 25979  
QY 1334 gLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrI 1354  
Db 25980 TCTGAATATGATCAAGCAGTGAAGAGTTAOCGCCAGCTGAAACAGGCAGCTATTACCAT 26039  
QY 1354 eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGl 1374  
Db 26040 TCAACCCGATATAGCCCAAGAGCAATGCAAAAGCAGGTGGTCTTGTACCAAAAGCA 26099  
QY 1374 nArgGluAlaIleIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGl 1394  
Db 26100 AAGGGAAGCCATTATCAAGTGCACGACGATACCGCGCATCTGGAGATGAGGAGCA 26159  
QY 1394 nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTTrpArgSe 1414  
Db 26160 GATTGAAGTCTACCAAAAAACAGCGCAGCTCATCCGCTTGCAGAAATGGTGGCGCAG 26219  
QY 1414 rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe 1434  
Db 26220 TATACCGCATGCGGCTGTGCAAGCGGGCTACCGAAGATTCGACTCAGTTCATTAG 26279  
QY 1434 rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe 1454  
Db 26280 CATTCACGCAAGTGGCGGCCACAGTGCAGGCTGTCGCCAACGGGAGATCTTCTTAAG 26339  
QY 1454 rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl 1474  
Db 26340 CACCATCCGCAAGTGGCACTTATGAGGCAITTTATCAGAGCAACTTTACTGATCGGTCA 26399

QY 1474 nGlnArgArgGluPheGluMetLysArgArgAlaAlaValValIleGlnArgArgPheAr 1494  
Db 26400 GCAGCGCAGGGATTCAGATGAAGGAGGGCTGCGGTAGTATCATCAGCGCGGTTTCG 26459  
QY 1494 gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerValI 1514  
Db 26460 TGTCTGCTGTCTATGCTTAAAGCGGAGACAGGAATACCAATTAATCCAAATCCTCTGTGAT 26519  
QY 1514 eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa 1534  
Db 26520 CCTGTGTCAGCGCAATTCGTGCTTAATCGCAGCATGAAGCAGCGCACGCCAGGAATTTGT 26579  
QY 1534 lGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetI 1554  
Db 26580 CCAGCTGCGCTACTATTTCGAGTCCATTTGCAACAAAAAGTTCCGTGGCAAGCGCTTAATGAT 26639  
QY 1554 eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl 1574  
Db 26640 TGAGCAGCGTAAATTGTTTCCAACTGCTCCGCTGTTCCATGCGCGGGCTTCCAGGACGTCG 26699  
QY 1574 aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe 1594  
Db 26700 CCGCGCTTTTATGGCTCGCAACAGGATTCCAGGCGCTGATGACACCCGAGATGATGACCT 26759  
QY 1594 uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrIleuLeAr 1614  
Db 26760 CATCGCCCAAGACGCGCGCCCAAGGTTATACAGCGTTTACTGCGCGGGCTTATCTAATCCG 26819  
QY 1614 gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGl 1634  
Db 26820 ACAGCTCAGAGACCAACGAGGTCTCTTGGATATCCGGAAGCGTATTGCTCAGCTCGGCA 26879  
QY 1634 nGluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuAr 1654  
Db 26880 AGAGGCAAAAGCGGTAAACTCTCTGCGCTGCAAGTCCAGGAGCGGTTCGTTCTTCG 26939  
QY 1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu----- 1669  
Db 26940 CCGAGCGCTTTATCGCATCAGATCGGTTAGCTAGTGTAACTGCGTGGTAAAGTATCCAG 26999  
QY 1670 -----AspArgLeuSerAr 1674  
Db 27000 CTACGTTAGTCTTATTATTAATACTGCTTAATAATATCCCGCAGATCGTCTTTCGCG 27059  
QY 1674 gThrValProHisLeuLeuMetTTrpCysSerGluPheMetSerThrPheCysTyrGlyI 1694  
Db 27060 CACTGTGCCACACCTGCTCATGTGTGTTTCGAGTTCATGTCCACGTTTTCGTATGGCAT 27119  
QY 1694 eMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgI 1714  
Db 27120 CATGCTCAGGCAATTCGATCAGAGTGGATAAGCAGCTTATAGAGCGCTGCAGCGGAT 27179  
QY 1714 eIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGluGlyGly 1734  
Db 27180 CATCTTAATTTGGCCCGCTACATAGCACACCGTGAACAGCTTCCAGAGCGCGGTTT 27239  
QY 1734 uValThrIleAlaGlnMetLeuLeuArgTTrpCysAspLysAspSerGluIlePheAsnTh 1754  
Db 27240 GGTCAACCATTTGCCAGATGTTTATTTCGCTGGTGGCGCACAAAGACAGTGAATTAACAC 27299  
QY 1754 rLeuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770  
Db 27300 TTTGTGCAACCTCATTTGGGTATTTCGCCCACTGCTCTAAAAAGCGAAAGTAGCTAAATA 27359  
QY 1771 -----IleIleH 1773  
Db 27360 CGAGAGGATTTAGGTTTATAGCGAGTTTAAATCATATTTCTTAATGCCCTTCCAGATCATTC 27419  
QY 1773 isAspTyrMetThrAsnProGluAlaIleTyrMetValArgGluThrLysLysLeuVal 1793  
Db 27420 ACGACTATGACCAACCCAGAGGCCATTTACATGTTGCGCGAAACTAAGAAAGCTCGTGG 27479

```

QY 1793 laArgIysGluLysMetLysGlnAlaAargLysProProMetThrSerGlyArgT 1813
Db 27480 CCGCAGGAAAGATGAAGCAAAATGACGCAAGCCGCCCAATGACAAAGTGACGGT 27539
QY 1813 YrLysSerGlnLysIleAsnPhenThrProCysSerLeuProSerLeuGluProAspPheG 1833
Db 27540 ATAAGAGCCAAAGATAAACTTCACGCGGTTCCTCCGCGCCAGGCTGAGCGGACITCG 27599
QY 1833 lylelleArgTyrSerProTyrThrPheleSerSerValTyrAlaPheAspThrIleL 1853
Db 27600 GAATCATCCGCTACAGTCCCTACAGTTTATCTCGTCCGTTTACGCCCTTCGATACGATT 27659
QY 1853 euCysLysLeuGlnIleAspMetPhe 1861
Db 27660 TGTCAAGCTGCAGATGCATGTTT 27685

RESULT 3
AC008205
LOCUS Drosophila melanogaster, 215711 bp DNA linear INV 24-FEB-2001
DEFINITION BACR33F18, complete sequence.
ACCESSION AC008335
VERSION AC008205.7 GI:13122716
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 215711)
Celiner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gokeyne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champ,M., Davenport,I.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,F., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ileguam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Furi,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 96A-96B
Unpublished
2 (bases 1 to 215711)
Celiner,S.E., Agbayani,A., Arcaina,T.T., Baxter,B., Blazej,R.G.,
Butenhoff,C., Champ,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 84-121, Berkeley, CA 94720, USA
On or before Feb 24, 2001 this sequence version replaced
gi:5701600, gi:7191021.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1..215711
/organism="Drosophila melanogaster"

```

```

/mol type="genomic DNA"
/strain="y; cn bw sp"
/db xref="taxon.7227"
/chromosome="3R"
/map="96A-96B"
/clone="BACR33F18 (D764)"
/clone lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 0 Length: 215711
Score: 9386.50 Matches: 1855
Percent Similarity: 96.32% Conservative: 4
Best Local Similarity: 96.11% Mismatches: 2
Query Match: 98.66% Indels: 69
DB: 3 Gaps: 3

US-09-914-698-1 (1-1861) x AC008205 (1-215711)

QY 1 MetCluLeuValTyrSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuIle 20
Db 40693 ATGGAGCTAGTGTGGAGCCCGCTCTTGAGGTGGCTTGCAGAGAGAGCGTGCAGCTAATA 40752
QY 21 AspAsnArgAsnPhenArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40
Db 40753 GACAAACGCACTTCCGAAGAGAGGTGATGATCATCTCAAGTCCAAGAGCAACGCGG 40812
QY 41 Val----- 41
Db 40813 GTCAAGGTCAGTCAAAAATATGCAATTCGCCACCCCTTTTACTTACATACCCCTCTTTTGCAT 40872
QY 42 -----LysAsnProArgLysPheProThrValGlyLysThrLe 54
Db 40873 GTATCTCTTACCTTTTACCTGAAGAAACCCGCGCAAAATTTCTTACTGTCCGCAAGACCT 40932
QY 54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValValSerAlaAlaVa 74
Db 40933 GCAGCTGAATCGCGAGAGAGCTGGCAAGACAATGAAAAGCGTGGTATCGCTGTGT 40992
QY 74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94
Db 40993 GCAGCAAAAGAGCGCATGTCTGCAGCAGCAGCGCTCCCTCCAAGCAGACATGCGGAGT 41052
QY 94 lThrAlaProSerArgProAlaAlaTyrAlaHisProProGlnAlaProLeuValG1 114
Db 41053 GACTGCTCTTTCGCGTCCCGCTGCATGGGCAATCCACCTCCAGAGCTCTCTTGTGCA 41112
QY 114 uLysAsnValTyrLysThrProGlnGluProValTyrIleSerProGlnProArgSe 134
Db 41113 GAAGNATGTATACAAGACTCCACAAGAAGAGCCCGGTATACATATCACCACAGCCTCGCAG 41172
QY 134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154
Db 41173 TCTTAAGGAAATCTAAGCCCTATGACGCCAGGAAACCTACTCGACGCTGATGCAATCT 41232
QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174
Db 41233 CGGATTACACCTCTCACCGAAACCCGTGGCAAGGACAGCTACCATTTTCCCGGACAA 41292
QY 174 nLeuAlaAlaTyrProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMe 194
Db 41293 TCTGGCAGCTGCGCCCAACCAACACATTAAGGGAATGAAATCATGTGCTAATGATAT 41352
QY 194 tArgProArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214
Db 41353 GCGGCGGTGCAATCACTCCAGATGATCTAGAGATCAGGTGCGCAACAAACGTT 41412
QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArg11 234
Db 41413 CGATGAAGCATTCGAGACTATCAATATTTGGTGGACACCTTGGGACTGTCTCCAGGAT 41472

```

FEATURES  
source



Db 43633 CCAGAGCAGCTATATCATGAGCTATCATAGTATTATCATCCATCCAGCGCTGTGCGGAGC 43692  
QY 974 aGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh 994  
Db 43693 CCAACAACCTGGGAAGGAGCAGCCGCGGGTGTGGAGCTCCGAGAGGCTGCGATCTT 43752  
QY 994 eLeuGlnArgIleThrArgArgGluPheAlaLysLysLeuLeuAlaAlaGluTh 1014  
Db 43753 TCTTCAGCGATCTGGCGGCGAGCTCTTTCGCCAAAACATATTGGCGCGCGGGAAC 43812  
QY 1014 rAlaArgLeuGlnArgSerGlnLysGlnAlaAlaAlaSerTyrIleGlnMetGlnTr 1034  
Db 43813 AGCCAGACTTCAGCGATGCCAAAACAACAGCAGCAGCTGTAGTATTATTCAAATGCAATG 43872  
QY 1034 pArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIl 1054  
Db 43873 GCGAAGCTATACGCTGGGAGAAATTCAGCGACAGGAGTTCTTCGCCGCGAGGAGCTCAT 43932  
QY 1054 eMetPheValGlnArgArgMetArgSerLysTyrTrpSerMetLeuGluGlnArgLysGluPh 1074  
Db 43933 CATGTTGTTTCAGCGCAGGATGCGAAGCAAGTGGAGTATGCTGGAGCAGCGCAAGAGTT 43992  
QY 1074 eGlnGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMe 1094  
Db 43993 CCAGCAACTAAAGCGTGCAGCTATATAATATCAACAACGCTGCGGAGCGAAGCTTTCAAT 44052  
QY 1094 tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAlaTy 1114  
Db 44053 GAGAAGTGCACGCTGATATTGTCGACTTCGTTCCAGCTTCCTTAAGTTTCAGGCTTA 44112  
QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
Db 44113 CAGGAAGCCACCAATCCAGATGAGAATAGATCGTAATCACTACTATTCCTCGCAAAAAA 44172  
QY 1134 nValIleCysLeuGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAs 1154  
Db 44173 TGTATCTCCCTGCACACGCTGAGGCCATCATGAAAAATGCGCAACAGAGGAAAA 44232  
QY 1154 nTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGl 1174  
Db 44233 TTATCTGAGGCTGCGAAATGCTTCAATACTAGTTCAAAAACCTACCGCATCGCTCAACA 44292  
QY 1174 nMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnAr 1194  
Db 44293 AATGATCCAGGATAGAAATGCAATATTAAGAACCCGCAATGATCATCAATGTCCAGAG 44352  
QY 1194 gArgTrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnTh 1214  
Db 44353 GCGCTGGAGGCCACTCTGCGAGATGCGTCGAGAAGGAGAACTACCTTCATCTCAAAAC 44412  
QY 1214 rThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl 1234  
Db 44413 GACAAACCAACGAATTCAAATCAAGTTCGTCGCCAAGCGCAATGAAAAAGCAAGAGC 44472  
QY 1234 aGluPheLeuGlnLeuLysValThrLeuValValGlnLysArgArgAlaLeuLe 1254  
Db 44473 CAGATTTCTTCAGCTGAAAAGGTTTACCTTGTAGTCCAGAAACCTCGCGGCGCTTTCCT 44532  
QY 1254 uGlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGl 1274  
Db 44533 GCAGATGCAAGAGGAGCGCAGAGTACCTACATCTGCGCGAGGTTTACATAAAGCTGCA 44592  
QY 1274 nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl 1294  
Db 44593 GCGCAGATTCAGATGCTCAAAAATCAATGCGGTTCATGCGAGCAAGTACCGCGGACCCA 44652  
QY 1294 nAlaAlaValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluAr 1314  
Db 44653 GGCCTGCTGTGAGCTGCTGCAGATGCAATTTGGGTTAATCATCTCTCTTAGGAAACCGGAGAG 44712  
QY 1314 gAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaAr 1334  
Db 44713 GAACAGTTTCTTACAACTCGCTCAAGCAGCAATAACATGCGACGACGATCGTCTTTTCGCG 44772

QY 1334 gLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIl 1354  
Db 44773 TCTGAATATGATCAAGCAGTTGAAGAGTTAGCCAGCTGAAACAGGACAGCTATTACCAT 44832  
QY 1354 eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGl 1374  
Db 44833 TCAAAACCCGATATAGACCCAGAGCAATGCAAAAGCAGGTGTCTTTGACCAAAAGCA 44892  
QY 1374 nArgGluAlaIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGl 1394  
Db 44893 AAGCGAAGCCATTATCAAGTGCAACGACGATACCGCGCAATCTCGAGATGAGGAAGCA 44952  
QY 1394 nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTrpArgSe 1414  
Db 44953 GATTGAAGTCTACCAAAAACAGCGCCAGGAGCTATCCGCTTCGCAAAATGGTGGCGCAG 45012  
QY 1414 rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe 1434  
Db 45013 TATACGCGACATGCGGCTGTGCAAGCGGCTACCGAAGGATTCGACTCAGTTCAATTAG 45072  
QY 1434 rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe 1454  
Db 45073 CATTCACGCAAGTGGCGGCCACAGTGCAGGCTCGTTCGCCAACCGGAGATCTTCTTAAG 45132  
QY 1454 rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl 1474  
Db 45133 CACCATCGCAAGTGGGACTTATGAGGACTTATCAGGCACTTTACTGATGGGTCA 45192  
QY 1474 nGlnArgArgGluPheGluMetLysArgArgAlaAlaValIleGlnArgArgPheAr 1494  
Db 45193 GCAGCGCAGGAAATTCAGATGAAGCAGAGGCTGCGTAGTGATCCAGCCGCGTTTCG 45252  
QY 1494 gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl 1514  
Db 45253 TGTCTGCTGTGCTATGCTAAAGCGGAGACAGGATTCACCAATTAATCCATTCCTCTGTGAT 45312  
QY 1514 eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa 1534  
Db 45313 CCTGCTCAGCGCAAAATTCGTCCTAATCGCAGCATGAGCAGGACGCCAGGAAATTTGT 45372  
QY 1534 lGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyIlyArgLeuMetIl 1554  
Db 45373 CCAGCTCGTACATTTGACAGTCCATTTGCAACAAAAAGTTCCGTCGCAAGCGTCTAATGAT 45432  
QY 1554 eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl 1574  
Db 45433 TGAGCAGCGTAATTTGTTTCCAACTGCTCCGCTCTTCATGCGCGGCTTCCAGCAGCTGC 45492  
QY 1574 aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe 1594  
Db 45493 CCGCGGCTTTATGGCTCGCAACGATTCAGGCGCTGATGACACCCGAGATGATGACCT 45552  
QY 1594 uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleAr 1614  
Db 45553 CATCCGCGCAAGCGCGCCCAAGGTTTATACAGCGTTACTGGCGGGGCTATCTAATCCG 45612  
QY 1614 gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGl 1634  
Db 45613 ACAGCGTCAGAAGCACACAGGCTCTCTTGGATATCCGGAAGCGGTATTGCTCAGCTGCGCA 45672  
QY 1634 nGluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuAr 1654  
Db 45673 AGAGGCAAGCGCGGTAAACCTCTGTCGCTGCAAAAGTCCAGGAGCGGTGCGTTCTTCGCG 45732  
QY 1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu----- 1669  
Db 45733 CGAGCGCTTTATCGCATCAGATCGGTAGCAGTGCTAAGTTCGATTCGATTCGGTAAAGTGATCCAG 45792  
QY 1670 -----AspArgLeuSerAr 1674  
Db 45793 CTAGTTAGTCTTATTATTAACTAATCTCTGCTTTAATAATCCCGCAGATCGTCTTTTCGCG 45852



QY	1674	gThrValProHisLeuLeuMetTyrCysSerGluPheMetSerThrPheCysTyrGly11	1694
DB	45853	CACGTGTCACACCTGCTCATGTGGTGTTCGGAGTTTCATGTCACAGCTTTTGTGTATGGCAT	45912
QY	1694	eMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArg11	1714
DB	45913	CATGGCTCAGCCCATTCGATCAGAGGTGGTATAGCAGCTTATAGAGCGCTGCGACCGGAT	45972
QY	1714	eIleuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGluGlyLe	1734
DB	45973	CATCTTAAATTTGGCCGCTACATACAGCACCGGTGAACAGCTTCAGAGGAGCGGT	46032
QY	1734	uValThrIleAlaGlnMetLeuLeuArgTyrCysAspLysAspSerGluIlePheAsnTh	1754
DB	46033	GGTCACCATTTGCCAGATGTTATTCGCTGCTGTCGACCAAGAGAGAGTGAATTCACAC	46092
QY	1754	rLeuCysThrLeuIleTyrValPheAlaHisCysProLysLysArgLys	1770
DB	46093	TTTGTGCACCTCTATTGGGTATTCGCCACTGTCTTAAAGAGCAAGTAGTCTAAATA	46152
QY	1771	-----IleIleH	1773
DB	46153	CGCAGAGGATTAGTGTTCAGCGAGTTTAAATCATATTTCTTAATGCCCTTCAGATCATTC	46212
QY	1773	isAspTyrMetThrAsnProGluAlaIleTyrMetValArgGluThrLysLeuValA	1793
DB	46213	ACGACTATATGACCAACCCAGAGCCCATTTACATGTGTCGCGAACTAAGAGCTCGTGG	46272
QY	1793	laArgLysGluLysMetLysGlnAsnAlaArgLysProProMetThrSerGlyArgT	1813
DB	46273	CCCCAAGGAAAGATGACGACAAATGACGACGACGCGCCCAATGACAAAGTGGACGGT	46332
QY	1813	yrLysSerGlnLysIleAsnPheThrProCysSerLeuProSerLeuGluProAspPheG	1833
DB	46333	ATAAGAGCCAAAGATAACTACGCGGTGTTCCTCCCGCAGCTGAGCCGCGACTTCG	46392
QY	1833	lyIleIleArgTyrSerProTyrThrPheIleSerValTyrAlaPheAspThrIleL	1853
DB	46393	GAATCATCCGCTACAGTCCCTACAGCTTATCTGCTCGTTTACGCTTCGATACGATTT	46452
QY	1853	euCysLysLeuGlnIleAspMetPhe	1861
DB	46453	TGTGCAAGCTGCAGATGCACATGTTT	46478
RESULT 4	AE003749	228433 bp DNA linear INV 14-FEB-2003	
LOCUS	AE003749	Drosophila melanogaster chromosome 3R, section 87 of 118 of the	
DEFINITION	AE003749	complete sequence.	
ACCESSION	AE003749.3	GI:23172203	
VERSION	AE003749.3	GI:23172203	
KEYWORDS			
SOURCE		Drosophila melanogaster (fruit fly)	
ORGANISM		Drosophila melanogaster	
REFERENCE		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
AUTHORS		1 (bases 1 to 228433)	
		Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,	
		Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Gallie,R.F.,	
		George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,	
		Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,	
		Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,	
		Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,	
		Abriel,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,	
		Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,	
		Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,	
		Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,	
		Burtis,K.C., Busan,D.A., Butler,H., Cadieu,E., Center,A.,	
		Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,	
		Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,	
		Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,	
		Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,	

REFERENCE		1 (bases 1 to 228433)	
AUTHORS		Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,	
JOURNAL		Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,	
REFERENCE		4 (bases 1 to 228433)	
AUTHORS		Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
JOURNAL		Submitted (13-FEB-2003) University of California Berkeley, 539 Life	
REFERENCE		6 (bases 1 to 228433)	
AUTHORS		Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,	
JOURNAL		Submitted (13-FEB-2003) University of California Berkeley, 539 Life	
REFERENCE		On Sep 18, 2002 this sequence version replaced gi:10726756.	
FEATURES		Location/Qualifiers	

TITLE		The genome sequence of Drosophila melanogaster	
JOURNAL		Science 287 (5461), 2185-2195 (2000)	
MEDLINE		20196006	
PUBMED		10731132	
AUTHORS		2 (bases 1 to 228433)	
		Celniker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A.,	
		Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y.,	
		Banson,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busan,D.A.,	
		Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,	
		Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,	
		Ferrera,S., Frise,E., Gallie,R.F., Garg,N.S., George,R.A.,	
		Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,	
		Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,	
		McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,	
		Paclet,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,	
		Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,	
		Stapleton,M., Strong,R., Swirskas,R., Tector,C., Tyler,D.,	
		Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.	
		Sequencing of Drosophila melanogaster genome	
		Unpublished	
		3 (bases 1 to 228433)	
		Mista,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,	
		Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochnik,S.E.,	
		Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W.,	
		Celniker,S.E., Champe,M.E., Drysdale,R.A., Emmert,D., Frise,E., de	
		Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B., Searle,S.M.J.,	
		Millburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E.,	
		Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,	
		Rubin,G.M., Mungall,C.J. and Lewis,S.E.	
		Annotation of Drosophila melanogaster genome	
		Unpublished	
		4 (bases 1 to 228433)	
		Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
		Direct Submission	
		Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,	
		Rockville, MD 20850, USA	
		5 (bases 1 to 228433)	
		Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.	
		Direct Submission	
		Submitted (06-SEP-2002) University of California Berkeley, 539 Life	
		Sciences Addition, Berkeley, CA 94720, USA	
		6 (bases 1 to 228433)	
		Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,	
		Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Gallie,R.F.,	
		George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,	
		Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,	
		Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,	
		Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,	
		Abriel,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,	
		Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,	
		Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S.,	
		Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,	
		Burtis,K.C., Busan,D.A., Butler,H., Cadieu,E., Center,A.,	
		Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,	
		Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,	
		Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,	
		Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,	





QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174  
DB 44803 GCGATTACACACCTCTACCGGAACCCGTGGCAAGAGCAAGACTACCAATTTCCCGGACAA 44862  
QY 174 nLeuAlaAlaTrpThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMe 194  
DB 44863 TCTGGCAGCCTGGCCACACCAACACTTAAAGGGAATGTAAATCATGTGCTAATGATAT 44922  
QY 194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214  
DB 44923 GCGCCCGCGTGGAACTACTCCAGATGATCTAGAGATCCGCTGCCCAACAAACGCTT 44982  
QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArgI 234  
DB 44983 CGATGTAAGACATTTCCGAGACTATCATATTTGGTGGACACCTTGACCTGCTCCAGGAT 45042  
QY 234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisThr 254  
DB 45043 CGATGGACAACCGCATACGCCCTTAAATAAGACAACAACCAATTTGTGCATGCCACGCACAC 45102  
QY 254 rArgAlaLeuAlaCysIleHisGluGluGlyProSerProArgThrProThrLys 274  
DB 45103 CAGAGCTCTGCGCTGTATTCATGAGGAGGAGGACCAAGTCCACTAGGACGCCCCACGAA 45162  
QY 274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294  
DB 45163 GAGCGCCATACACGACTGAAGAGGACANTAGTTGGTGGTTTACCCCTTACGAAGTA 45222  
QY 294 rSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyzAlaIleGlnG 314  
DB 45223 TTTCCGAGTCCATGAAAGATTTGTCACTTCATCGCCACAACTAAGTATGCCATTCAAAG 45282  
QY 314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyzrGl 334  
DB 45283 GTCTATGCGCTTAATCTAAATGAATGAATCCCGCTCGATCGAACAAGATAGATATTACCA 45342  
QY 334 nGluGlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaIse 354  
DB 45343 GGAGCAGCAGATCCAGATAAAGCCCAAGACTTGATAGCTCTCTAGTAGCGAGGCTAG 45402  
QY 354 rLeuAlaGlyGlnGlnPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374  
DB 45403 TTTGGCGGCGCAGCAGAGTTTCTAATCAACACACAGTGAGATCCTCGCTCAGTCCAGTGC 45462  
QY 374 gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi 394  
DB 45463 TTTTAACTCCATGAAGTAGGTGGGAAGTCGGTGAAGGGAAGTCCGGTGAAGGAATCCTCA 45522  
QY 394 sLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyzrAr 414  
DB 45523 CAAGCGCGCTCTCATGAGTTGAGTTTTCGGATGCACTAGCAAGATCATTTGTACCG 45582  
QY 414 gAsnGluThrValAlaIleSerProLysLysGlnArgValGluAspThrThrLeuPr 434  
DB 45583 CAATGAATCTGAGCATTTCCCTCTCTAAAGAACCAACGGGTTGAGACACTACTCTGCC 45642  
QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHi 454  
DB 45643 CAGAAGTGAGCGCGCGGCAAAATGCATCTCAGAAAGCAGTAGTGGCCACCGCTGGCCACA 45702  
QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
DB 45703 CGCCCAATCCAAGAAGTTAAGCTAGCACAAACCATTTGTCTCATGTATGAAGAGCCGCCAC 45762  
QY 474 rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyzrAspSerGlu 494  
DB 45763 ACCACGAAAAGTCAGGACACTAGCATTCAGCCTTCGGTCAAGCTCTATGACTCGGAGCT 45822  
QY 494 uTyzrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThrIleAspProPh 514  
DB 45823 GTATATGCAGAGTGTCATCAACCCCGATCCATTTGAGCAACTTACGACAAATGATCCATT 45882

QY 514 eLeuAlaSerThrMetTyzrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy 534  
DB 45883 TCTGGCATCTACCATGATTTTGGATGAACAGGCTGTGGATCGTCATCAAGCTGACTTTAA 45942  
QY 534 sLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnLys 554  
DB 45943 AAAGTGGTTAAATGCCCTTCTCTCCATACCGCTGACCTGGACCGCAGATTTAATAACAA 46002  
QY 554 sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValAlaLProTh 574  
DB 46003 AATAGACGTTGGTAAGCTGTTTAAACGAGGTGCGCAACAAAGAGCTCGTGGTGGCTCCAC 46062  
QY 574 rLysGluGluGlnSerMetAsnTyzrLeuThrLysTyzrArgLeuGluThrLeuArgLysAl 594  
DB 46063 CAAGGAGGACAGCTCTATGAACTACCTTACGAATATCCGCTCGAGACGCTTGTAAAGGC 46122  
QY 594 aAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTy 614  
DB 46123 GGCTGTGGAGCTCTTCTTTCAGTGAGCAGATGCGCTGCCATGCTCCAAAGTGGCCGTATA 46182  
QY 614 rValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMe 634  
DB 46183 TGTCAACAAGCAAGCTCTGCGCATCCGTAGCGATCGTAATCTTCACTAGACGTAGTTAT 46242  
QY 634 tGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLe 654  
DB 46243 GCNAACGACCATACTTGGAGCTGCTGCTTCTTCAATCCCTTTGGTGGCTCGCCCTGGACT 46302  
QY 654 uGluValValPheGlyGlyLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe 674  
DB 46303 GGAAGTGGTCTTTGGCGAGAGATCCAGATGCACTCTAATCGAGACATTTAGTGGCTCCAG 46362  
QY 674 rThrPheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyzrSerLysAl 694  
DB 46363 CACCTTTATCTCAATCGCTTGTTCGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC 46422  
QY 694 aTyzrThrLeuThrGluGluTyzrAlaGluThrIleLysLysHisSerLeuGlnLysIleLe 714  
DB 46423 ATACACACTCACGAAGAGTACGCGAGACCATTAAGAAGCAGCTCATTCAGAAAAATCCT 46482  
QY 714 uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCy 734  
DB 46483 CTTTTTGTCTTCTTCGATCAAGCTAAGCAGAGCGCATCGTCAAGCACATTCCTTG 46542  
QY 734 sLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSe 754  
DB 46543 TTTGTTTGTAAAGTGCCTCCACATTAAGAGACCAAGGATATTCGCTCGCTTCTCGTC 46602  
QY 754 rGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyzrValle 774  
DB 46603 GGAGCTGCTCGCAACATTTGGTGATATTACGCGGGAACCTTCGTCGCTCGGCTTACGTTCT 46662  
QY 774 uGlnHisArgGlnThrPheLeuAspGluPheAspTyzrAlaPheAsnAsnLeuAlaValAs 794  
DB 46663 ACAGCACCGCCCAACATTTTGGACGAGTTGATTTATGCTTCAACAACTTGGCTGGA 46722  
QY 794 pLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspAsple 814  
DB 46723 CCTAAGAGATGGCGTGAGACTAACCGGAGTATGAGGTAATTTTACTACGCGCATGATCT 46782  
QY 814 uThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLe 834  
DB 46783 AACCGCCCATTAAGGTCGCGCATCTCTCGCTTACGCGGATCTTCAATGTAAAGCT 46842  
QY 834 uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAs 854  
DB 46843 AGCTCTGGGCGCACCTTGTGAAGCCCACTTCCAGCTAGCGCGGACATCGCGCCCAAGA 46902  
QY 854 pIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIleTyzrLysPh 874  
DB 46903 CATCGTTGACGAGACATCGTGAGAGACGCTTTCCCTGCTCTGGCACTTATTATTACAAAT 46962  
QY 874 eArgSerProLysPheHisAlaAlaThrValLeuGlnLysTyzrTrpArgArgHisTr 894

Db	46963		CCGCTCGCCCAAGTTTCATGCGCGGCCACCGTGCTCCAGAAATGGTGGCGCGTCACTG	47022	Db	48043	CGAGTTTCTTCAGCTGAAAAAGGTTACCTCTGTAGTCCAGAAACGTCGGCGGGCTTTGGCT	48102			
Qy	894	pLeuHisValValIleGlnArgArgIleArgHisLysGlnLeuMetArgArgHisArgAl	914	Qy	1254	uGlnMetArgLysGlnuArgGlnGluTyrLeuHisLeuArgGlnuValThrIleLysLeuGl	1274	Qy	1254	uGlnMetArgLysGlnuArgGlnGluTyrLeuHisLeuArgGlnuValThrIleLysLeuGl	1274
Db	47023	GCTGCACGTTGTTATTCAGCGCTGCATTCGCCACAAAGAGCTTATGCGTCGCCACCGGGC	47082	Db	48103	GCAGATCGAAAGAGGAGCGCCAGAGTACCTACATCTCGCGAGGTTACCATATAAGCTGCA	48162	Db	48103	GCAGATCGAAAGAGGAGCGCCAGAGTACCTACATCTCGCGAGGTTACCATATAAGCTGCA	48162
Qy	914	aAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPh	934	Qy	1274	nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl	1294	Qy	1274	nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl	1294
Db	47083	CGCTACTGTCTATTGAGGCCGTTTCGCTGGCCACAGATGAGAAAGTACGTGAAGTTGTT	47142	Db	48163	CGCGAGATTCATGCTCAAAAATCAATCGCGTTTCATCGAGCCCAAGTACCGCGCACCCA	48222	Db	48163	CGCGAGATTCATGCTCAAAAATCAATCGCGTTTCATCGAGCCCAAGTACCGCGCACCCA	48222
Qy	934	eLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAl	954	Qy	1294	nAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluAr	1314	Qy	1294	nAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluAr	1314
Db	47143	TAAGCGGAAGCACTCAGCGCGCAATTAATCTGCAGAAAGTTCACCCGTCGTTATTTGGC	47202	Db	48223	GGCTGTCTGTAGCTGCGCTGCAGATGCATTGGCGTAATCATCTCCTTAGGAACCGGAGAG	48282	Db	48223	GGCTGTCTGTAGCTGCGCTGCAGATGCATTGGCGTAATCATCTCCTTAGGAACCGGAGAG	48282
Qy	954	aGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTyrTrpArgAl	974	Qy	1314	gAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaAr	1334	Qy	1314	gAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaAr	1334
Db	47203	CCAGAAAGCAGCTATATCAGAGCTATCAGATATATCACCATCCAGCGCTGCTGGCGAGC	47262	Db	48343	GAAACAGTTTCTTCAACTGCGTCAAGCGCAATAACTGCGAGCGACGGTACCGAGCTCG	48342	Db	48343	GAAACAGTTTCTTCAACTGCGTCAAGCGCAATAACTGCGAGCGACGGTACCGAGCTCG	48342
Qy	974	aGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePh	994	Qy	1334	gLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIl	1354	Qy	1334	gLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIl	1354
Db	47263	CCAAACAATGGGAAGGACGACCGCCAGCGGTTTGTGGAGCTCCGAGAGGCTGCGATCTT	47322	Db	48343	TCTGAATATGATCAAGCAGTTTGAAGAGTTACGCCAGCTGAAACAGGCGAGCTATTACCAT	48402	Db	48343	TCTGAATATGATCAAGCAGTTTGAAGAGTTACGCCAGCTGAAACAGGCGAGCTATTACCAT	48402
Qy	994	eLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluTh	1014	Qy	1354	eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValLysLeuTyrGlnLysGl	1374	Qy	1354	eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValLysLeuTyrGlnLysGl	1374
Db	47323	TCTTCAGCGGATCTGGCGGCGACGACTCTTTGCCAAAAAACTATTGGCGCGCGGAAAC	47382	Db	48403	TCAAAACCGATATAGAGCCAAAGAGCAATGCAAAAGCAGGTGGTCTTTGTACCAAAAGCA	48462	Db	48403	TCAAAACCGATATAGAGCCAAAGAGCAATGCAAAAGCAGGTGGTCTTTGTACCAAAAGCA	48462
Qy	1014	rAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTr	1034	Qy	1374	nArgGluAlaIleLysLysValGlnArgArgTyrArgGlyValAsnLeuGluMetArgLysGl	1394	Qy	1374	nArgGluAlaIleLysLysValGlnArgArgTyrArgGlyValAsnLeuGluMetArgLysGl	1394
Db	47383	AGCCAGACTTCAGCGATCGCAAAAAACAACAGCAGCTGCTAGTTATATTCAAATGCAATG	47442	Db	48463	AAGGAGAGCCATTTCAAGTGCAACGACGATACCGCGGCAATCTGGAGATGAGGAAGCA	48522	Db	48463	AAGGAGAGCCATTTCAAGTGCAACGACGATACCGCGGCAATCTGGAGATGAGGAAGCA	48522
Qy	1034	pArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIl	1054	Qy	1394	nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrTrpArgSe	1414	Qy	1394	nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrTrpArgSe	1414
Db	47443	CGGAAGCTATCAGCTGGGCGAATTCAGCAGCGTGTTCTGCGCCAGAGGGACCTCAT	47502	Db	48523	GATTGAAGCTTACCAAAAACAGCGCCAGCGCATCTCCGCTTGCAAAAATGGTGGCGAG	48582	Db	48523	GATTGAAGCTTACCAAAAACAGCGCCAGCGCATCTCCGCTTGCAAAAATGGTGGCGAG	48582
Qy	1054	eMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPh	1074	Qy	1414	rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe	1434	Qy	1414	rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe	1434
Db	47503	CATGTTTGTTCAGCGCAGATCGGAAGCAAGTGGAGTATGCTGGAGCAGCGCAAGAGTT	47562	Db	48583	TATACGCCACATCGCGGCTGTCAAAGCGGGCTACCGAAGGATTCGATCAGTTCATTGAG	48642	Db	48583	TATACGCCACATCGCGGCTGTCAAAGCGGGCTACCGAAGGATTCGATCAGTTCATTGAG	48642
Qy	1074	eGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMe	1094	Qy	1434	rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe	1454	Qy	1434	rIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe	1454
Db	47563	CCAGCAACTAAAGCGTGCAGCTATAAATATCCAAACGCTGGCGAGCAAGCTTTCAAT	47622	Db	48643	CATTCAACGCAAGTGGCGGGCCACAGTGCAGGCTGTCGCCACGCGGAGATCTTCTTAAG	48702	Db	48643	CATTCAACGCAAGTGGCGGGCCACAGTGCAGGCTGTCGCCACGCGGAGATCTTCTTAAG	48702
Qy	1094	tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTy	1114	Qy	1454	rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl	1474	Qy	1454	rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl	1474
Db	47623	GAGAAAGTGCACCGCTGATATTTCGCACTTCGTTCCAGCGTTCITAAAGTTACGCTTA	47682	Db	48703	CACCATCGCAAAAGTGCAGCTTATGCGAGCATTTATCAGAGCAACTTTACTGATCGGTCA	48762	Db	48703	CACCATCGCAAAAGTGCAGCTTATGCGAGCATTTATCAGAGCAACTTTACTGATCGGTCA	48762
Qy	1114	rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs	1134	Qy	1474	nGlnArgArgGluPheGluMetLysArgArgAlaAlaValIleGlnArgArgPheAr	1494	Qy	1474	nGlnArgArgGluPheGluMetLysArgArgAlaAlaValIleGlnArgArgPheAr	1494
Db	47683	CAGGAAGCCCAATCCAGATGAGATAGATCGTAATCACTACTATTTCCTCGCAAAAAA	47742	Db	48763	GCAGCGCAGGGAATTCGAGATGAAGCGAAGGGCTGCCGTAGTATCCAGCGCGGTTTCG	48822	Db	48763	GCAGCGCAGGGAATTCGAGATGAAGCGAAGGGCTGCCGTAGTATCCAGCGCGGTTTCG	48822
Qy	1134	nValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAs	1154	Qy	1494	gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl	1514	Qy	1494	gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl	1514
Db	47743	TGTTATCTGCTGCAACAGCACTGAGGCGCATCATGAAATGCGCAACAGAGGGAATA	47802	Qy	48823	TGCTCGCTGTGCTATGCTTAAGGCGAGACAGGATTAACCAATTAATCCATCTCTGTAT	48882	Qy	48823	TGCTCGCTGTGCTATGCTTAAGGCGAGACAGGATTAACCAATTAATCCATCTCTGTAT	48882
Qy	1154	nTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGl	1174	Qy	1514	eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa	1534	Qy	1514	eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa	1534
Db	47803	TTATCTGAGCGTGCAGAAATGCTTCAATAGTTCAAAAACGCTACCGCATCGCTCAACA	47862	Db	48883	CCTGTGTGCGCGCAAAATTCGCTGCTAAATCGCAGCATGAAGCAGGCGACGCGAGAAATTTGT	48942	Db	48883	CCTGTGTGCGCGCAAAATTCGCTGCTAAATCGCAGCATGAAGCAGGCGACGCGAGAAATTTGT	48942
Qy	1174	nMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnAr	1194	Qy	1534	lGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIl	1554	Qy	1534	lGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIl	1554
Db	47863	AATGATCCAGGATAGAAATGCATATTTAAGAACCCGCAATGTATCATCAATGTCAGAG	47922	Qy	48943	CCAGCTGGGTACTATTGTGAGTCCATTTTGCCAATAAAGTTCCGTCGCGACGCTTAATGAT	49002	Qy	48943	CCAGCTGGGTACTATTGTGAGTCCATTTTGCCAATAAAGTTCCGTCGCGACGCTTAATGAT	49002
Qy	1194	gArgTrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnTh	1214	Qy	1554	eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl	1574	Qy	1554	eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl	1574
Db	47923	CGCGTCGAGAGCCACTCTGCAGATCGTGCAGAAAGAAAGAACTACTTTCATCTCCAAAC	47982	Qy	49003	TGAGCAGCGTAATGCTTTTCCAACTGCTCGCTGTTCCATGCGCGGCTTCCAGGCGACGTCG	49062	Qy	49003	TGAGCAGCGTAATGCTTTTCCAACTGCTCGCTGTTCCATGCGCGGCTTCCAGGCGACGTCG	49062
Qy	1214	rThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAl	1234	Qy	1574	aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe	1594	Qy	1574	aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe	1594
Db	47983	GACAACCAACCAATTCAAATCAAGTTTCGTGCCAAGCGCGAAATGAAAAAGCAAGAGC	48042	Qy	49063	CGCGCGCTTATGGTTCGCAACAGATTCAGGCGCTGATGACACCGCAATGATGACCTT	49122	Qy	49063	CGCGCGCTTATGGTTCGCAACAGATTCAGGCGCTGATGACACCGCAATGATGACCTT	49122
Qy	1234	aGluPheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLe	1254	Qy	1594	uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleAr	1614	Qy	1594	uIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleAr	1614

```

1614 gAArgGlnGlyHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgG1 1634
49183 ACAGAGTCAGAGACACAGAGGTCCTCTGGATATCCGGAAGCGTATTGCTCAGTGGCGCA 49242
1634 nGUaLaLaValaValaSerValArgCysLysValGlnGluLaValaValArgPheLeuAr 1654
49243 AGAGGCAAGAGCGGTAACTCTGGCGCTGCAAGTCCAGAGGCGGTGGTTCCTTCG 49302
1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu----- 1669
49303 CGAGCGCTTTATCGCATCAGATCGGTAGCAGTGAAGTTCGATGGTAAAGTGCATCCAG 49362
1670 -----AspArgLeuSerAr 1674
49363 CTACGTTAGTCCTTATTATAACTAATCTGCTTAATAATCCCGGAGATCGCTTCGCG 49422
1674 gThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTrpGlyI1 1694
49423 CACTGTGCCACACCTGCTCATGTGTGTGTTTCGGAGTTCATGTCCACGTTTGTCTATGCCAT 49482
1694 eMetaLagInAlaIleArgSerGluValaAspLysGlnLeuIleGluArgCysSerArgI1 1714
49483 CATGGCTCAGGCCATTCGATCAGAGGTGGAATAAGCAGCTTATAGAGCGCTGCAGCCGGAT 49542
1714 eIleLeuAsnLeuAlaArgTrpAsnSerThrThrValAsnThrPheGlnGluGlyGlyLe 1734
49543 CATCCTAAATTTGGCCGCTACATAGCACCCAGCGGTGACACGTTCCAGAGGCGCGTTT 49602
1734 uValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluIlePheAsnTh 1754
49603 GGTCCACCATTTGCCAGATGTTATTGGCTGTGGTGCACAAAGACAGTGCATATTCACAC 49662
1754 rLeuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770
49663 TTTGTGCACCTTCATTTGGTATTTCGCCACTGTCCTAAAGACGGAAGAGTAGTAAATA 49722
1771 -----IleIleH 1773
49723 GCGAGAGGATTAGTTTACGAGTTTAAATCATATCTTAAATGCTTCGATCATTC 49782
1773 iAspTrpMetThrAsnProGluAlaIleTyMetValArgGluThrLysLysLeuVala 1793
49783 AGCACTATATGACCAACCCAGAGGCGCATTTACATGTGGCGGAAACTAAGAAAGCTCGTG 49842
1793 laArgLysGluLysMetLysGlnAsnAlaArgLysProProMetThrSerGlyArgT 1813
49843 CCGCAAGGAAAGATGAGCAAAATGACGCAAGCCGCGCAATGACAAAGTGCAGCGT 49902
1813 yrLysSerGlnLysIleAsnPheThrProCysSerLeuProSerLeuGluProAspPheG 1833
49903 ATAGAGCCAAAGATAAACTTCACGCGGTGTTCCCTGCCAGCTGCAGCGGACTTCG 49962
1833 lyIleIleArgTrpSerProTyThrPheIleSerSerValTyAlaPheAspThrIleL 1853
49963 GAATCATCGCTACAGTCCCTACAGTTTATCTCGCTTCGTTTACGCCCTCGATACGATT 50022
1853 euCysLysLeuGlnIleAspMetPhe 1861
50023 TGTGCAAGCTGCAGATCGACATGTTT 50048

```

```

RESULT 5
AY094825
LOCUS Drosophila melanogaster LD39479 full insert cDNA.
DEFINITION Drosophila melanogaster (fruit fly)
ACCESSION AY094825
VERSION AY094825.1 GI:20151636
KEYWORDS FLI CDNA.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

```

## REFERENCE

## AUTHORS

1 (bases 1 to 3486)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,  
 Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,  
 George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G.,  
 Miranda, A., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,  
 Patel, S., Phouanavong, S., Wan, X., Yu, C., Lewis, S. E., Rubin, G. M.  
 and Celniker, S.

## Direct Submission

Submitted (03-APR-2002) Berkeley Drosophila Genome Project,  
 Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, USA

## Sequence submitted by:

Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to  
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
 Science 2000). The sequence has been subjected to integrity checks  
 for sequence accuracy, presence of a polyA tail and contiguity  
 within 100 kb in the genome. Thus we believe the sequence to  
 reflect accurately this particular cDNA clone. However, there are  
 artifacts associated with the generation of cDNA clones that may  
 have not been detected in our initial analyses such as internal  
 priming, priming from contaminating genomic DNA, retained introns  
 due to reverse transcription of unspliced precursor RNAs, and  
 reverse transcriptase errors that result in single base changes.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our Web site  
 (<http://fruitfly.berkeley.edu>) or send email to  
[cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

## Location/Qualifiers

```

1..3486
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="96A23-96A23"
1..3486
/gene="asp"
/note="alignment with genomic scaffold AE003749"
/db_xref="FLYBASE:FBgn0000140"
82..2715
/gene="asp"
/note="Longest ORF"
/codon_start=1
/product="LD39479p"
/protein_id="AAM11178.1"
/db_xref="GI:20151637"
/db_xref="FLYBASE:FBgn0000140"
/translation="MORTILSELLCFNPLRLGLLEVVFVFGKIQMSNRDIVGLSTFI
LNRFRNKCEQRYSKAYLTETVETATIKHSLOKILFLLLDQAKQKRIYKHNPL
FVKSPKBTEDILLRPSSELLANGIDITRELRILGYLQHQTPLEDFVAFNNLAV
LDRGPRLTRVMEVILLDDTLRLRVPAIQRIFNVKLALGALGANFALQGDIA
AODIVDGHREKTLISLLWLIYKFRSPKFAAATVLQKWRHRLHVLVQRIRKHELM
RRHAATVIOAFVGHGOMKVKLFTKRTQAAIILQKFTRYLAKOLYQSYHSIIT
IQWRRAQLQGRHQRFVEUREAIFLQIRWRLFAKLAAAEATLQRSQQA
AASTIQMSRYSQIGRIQRFQRLDIFVQRMESKMSLQKKEFQOLKRAAIN
IQWRRAKLINRKNADYLSLVKYAYKATIQMRIDRNHYSLRKVCIQOR
LRATMKREQRNENILRLNLSILVKRYRQMIQDRNAYLRTKCIINVRWRAT
LOMRERKNVLIQTTTKRIQIKFAKREKKORAEFLQAKVILVQKRRLQMR
KERQYLHLREVTIKLQRRFQAQSMRPMARKYGTQAAVSCLOHWRNHLRKRRN
SFLQROAATILQRRLNMLIKOLSYAOLKQAATITQYIRAKKAMQKVLYOK
QREAITVQRRYRGNLEMRKQIEVYQKQAVTLQKWRIRDMRLCKAGYRIRLS
SLSTQKRWATVQARQREIFLSTIRKRLMQAIFRCVCSAKSIGSSFAHCATPAHVV
FGVHVHLLMHGSHGSHIRGG"

```

## CDS

## gene

## ORIGIN

Alignment Scores:  
 Pred. No.: 4.22e-287 Length: 3486  
 Score: 5267.00 Matches: 1055  
 Percent Similarity: 84.32% Conservative: 4  
 Best Local Similarity: 84.00% Mismatches: 2  
 Query Match: 55.36% Indels: 195

DB: 3 Gaps: 1

US-09-914-698-1 (1-1861) x AY094825 (1-3486)

607 ProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgLysArgSerAspArg 626

DB 1 CCATGCTCCAAAGTGGCGGTATATGCAACAAGCAAGCTCTGCGCATCCGTAGCGATCGT 60

627 AsnLeuHisLeuAspValValMetGlnArgThrLleLeuGluLeuLeuLeuCysPheAsn 646

DB 61 AATCTTCACCTAGACGTAGTATATGCAACGACCATCACTTGAGCTGCTGCTTCAAT 120

647 ProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysLleGlnMetGlnSer 666

DB 121 CCCCTTTGGCTGCGCCCTTGACGTGAAGTGGTCTTTGGCGAAGATCCAGATGCACTCT 180

667 AsnArgAspLleValGlyLeuSerThrPheLleLeuAsnArgLeuPheArgAsnLysCys 686

DB 181 AATCGAGACATGTTAGGCCTCAGCACCTTTATCCTCAATCGCTTGTTCGGAATAAGTGT 240

687 GluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTyrAlaGluThrLleLys 706

DB 241 GAGGAGCAGAGGATACAGCAAGGCATACACATCCAGGAAGTACGCGGAGACCATTAAG 300

707 LysHisSerLeuGlnLysLleLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLys 726

DB 301 AAGCAGCTCATTCAGAAAATCCTCTTTTGTGCTTTTCTCGATCAAGCTAAGCAAG 360

727 ArgLleValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLys 746

DB 361 CGCATCTCAAGCAACATCCCTGTTGTTGTTAAAGTCCGCACATAAAGAGACCAAG 420

747 AspLleLeuLeuArgPheSerSerGluLeuLeuAlaAsnLleGlyAspLleThrArgGlu 766

DB 421 GATATCTGCTGCGCTCTCTCGTCGAGCTGCTCGCAACATGTGTGATTAATACCGGGA 480

767 LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr 786

DB 481 CTTCGTGCGCTGGCTAGCTCTACAGCACCGCCAAACATTTTGGACGAGCTCGATAT 540

787 AlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGlu 806

DB 541 GCCTTCAACAACTTGGCTGTGACCTAAGATGGCGTGAGACTAAACCGAGTCATGGAG 600

807 ValLleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaLleSerArgLeu 826

DB 601 GTAATTTTACTACGCGATGATCTAACCGCCAGTTAAGGGTGCCAGCCATCTCTCGCCT 660

827 GlnArgLlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu 846

DB 661 CAGCGGATCTCAATGTAAAGCTAGCTCTGGGCGCACTTGTGTGAAGCCCACTTCCAGCTA 720

847 GlyGlyAspLleAlaAlaGlnAspLleValAspGlyHisArgGluLysThrLeuSerLeu 866

DB 721 GCGCGGACATCGCGCCCAAGACATCGTTGACGGACATCGTGAGAAGACGCTTTCCCTG 780

867 LeuTrpGlnLeuLleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeu 886

DB 781 CTCGTGGCAACTTATTTTACAAATTCGCTGCCCAAGTTTCAATGCGCGCCGACGGTGTCT 840

887 GlnLysTrpTrpArgArgHisTrpLeuHisValValLleGlnArgArgLleArgHisLys 906

DB 841 CAGAAATGTTGGCGCGCTCACTGGCTGCAAGTGTGTTATCCAGCGTCGATTCGCCACAAA 900

907 GluLeuMetArgArgHisArgAlaAlaThrValLleGlnAlaValPheArgGlyHisGln 926

DB 901 GAGCTTATCGTCGCGCACCGGCCCTACTGTTCATTCAGGCCGTGTTCGTGGCCACCGAG 960

927 MetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaLleLleLeuGln 946

DB 961 ARGAGAAATAGCTGTAAGTTGTTTAAAGCGGAACGACATCAGCGCCGCAATATTTCTGCA 1020

947 LysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerLleLle 966

DB 1021 AAGTTCCACCGCTCGGTATTTGGCCAGAACAGCTATATCAGAGCTATCAGATTTATC 1080

QY ThrLleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVal 986

DB 1081 ACCATCCAGCGCTGGTGGCGAGCCCAACAACCTGGGAAGGAGCAGCACCGCCACGCGTTGTG 1140

QY GluLeuArgGluAlaAlaLlePheLeuGlnArgLleTrpArgArgArgLeuPheAlaLys 1006

DB 1141 GAGCTCCGAGAGGCTCGATCTTTCTTCAGCGGATCTGGCGGACGACCTCTTTGCCAAA 1200

QY LysLeuLeuAlaAlaGluThrAlaArgGlnArgSerGlnLysGlnGlnAlaAla 1026

DB 1201 AAACTATTGGCGCGCGGAAACAGCCAGACTTCAGCGATCGCAAAAACACAGCGAGCT 1260

QY AlaserTyrLleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgLleGlnArgHisGlu 1046

DB 1261 GCTAGTTATATTAATGCAATGGCGAAGCTATCAGCTGGCGAGAAATTCAGCGACAGCAG 1320

QY PheLeuArgGlnArgAspLeuLleMetPheValGlnArgArgMetArgSerLysTrpSer 1066

DB 1321 TTCTGCGCCAGAGGAGCTCATCATGTTTTCAGCGCAGGATCGGAAGCAAGTGGAGT 1380

QY MetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaLleAsnLleGlnGln 1086

DB 1381 ATGCTGAGCAGCGCAAGGAGTTCCAGCAACTAAGCGTGCGAGCTATAAATATCCAAACA 1440

QY ArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSer 1106

DB 1441 CGCTGGCGAGGAAGCTTCAATGAGAAAGTGCAACGCTGATTTATTTGGCACCTCGTTC 1500

QY SerValLeuLysValGlnAlaTyrArgLysAlaThrLleGlnMetArgLleAspArgAsn 1126

DB 1501 AGCGTTCTTAAAGTTTCAGGCTTACAGGAAAGCCCAATCCAGATCAGATAGATCGTAAT 1560

QY HisTyrTyrSerLeuArgLysAsnValLleCysLeuGlnGlnArgLeuArgAlaLleMet 1146

DB 1561 CACTACTATTCCCTGCGAAAATAATGTTATCTGCTGCAACAGCGACTGAGGGCCATCATG 1620

QY LysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerLleLeuValGln 1166

DB 1621 AAAATGCGCAACAGAGGGAATAATCTGAGGCTGCGAATGCTTCAATACTAGTTCAA 1680

QY LysArgTyrArgMetArgGlnGlnMetLleGlnAspArgAsnAlaTyrLeuArgThrArg 1186

DB 1681 AAACGCTACCGCATGCGTCAACAAATGATCCAGATAGAAATGCAATTTTAAAGACCCGC 1740

QY LysCysLleLleAsnValGlnArgArgTrpArgAlaThrLeuGlnMetArgArgGluArg 1206

DB 1741 AAATGTATCATCAATGTCAGAGCGCTGGAGAGCCACTCTGCAGATCGCTCGAAGAGG 1800

QY LysAsnTyrLeuHisLeuGlnThrThrLysArgLleGlnLleLysPheArgAlaLys 1226

DB 1801 AAGAACTACTTTCATCTCAAAACGCAACCAACGAATTCAAATCAAGTTCGCTGCCAAG 1860

QY ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVal 1246

DB 1861 CGCGAAATGAAAAAGCAAGCGCGAGTTCTTCAGCTGAAAAAGGTTTACCTTTGATGTC 1920

QY GlnLysArgArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyrLeuHisLeu 1266

DB 1921 CAGAAACCTCGCGCGCTTTCGTGAGATGCGAAGAGGAGCGCCAGAGTACTACATCTG 1980

QY ArgGluValThrLleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMet 1286

DB 1981 CGCGAGGTACCAATAAGCTCAGCGCAGATTCATGCTCAAAAATCAATGCGGTTCAAG 2040

QY ArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsn 1306

DB 2041 CGAGCAAGTACCGCGCACCCAGCGCTGCTGTGAGCTGCTGCAGATGATGATGGCGTAAT 2100

QY HisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaLleThr 1326

Db 2101 CATCTCTTAGGAAACGGGAGGAAACAGATTCTTCAACCTGCTCAAGCAGCAATAACA 2160

QY 1327 LeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnIleLysSerTyrAlaGln 1346

Db 2161 CTGCAGGACGGTACCGAGCTCGTCTGAATATGATCAAGCAGTTGAAGAGTTAGCCCGAG 2220

QY 1347 LeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLys 1366

Db 2221 CTGAAAACAGCAGCTATTACCATTTCAAAACCGCATATAGACCAAGAGGCAATGCAAAAG 2280

QY 1367 GlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArgArgTyrArg 1386

Db 2281 CAGGTGCTCTTGACCAAAAACCAAGGAGGAGCCATTATCAAGTGCACACGATACCGC 2340

QY 1387 GlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIle 1406

Db 2341 GGCATCTGGAGATGAGGAGCAGATTGAAGTCTACCAAAACAGCCCGAGGAGTCATC 2400

QY 1407 ArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArg 1426

Db 2401 CGCTTGCAAGAAATGGTGGCGCAGTATACGCGCATCGGCTGTGCAAGCGGCTACCGA 2460

QY 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArg 1446

Db 2461 AGGATTGCACTCGATTCATTGAGCATTTCAACGCAAGTGGCGGCGCACAGTGCAGGCTCGT 2520

QY 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466

Db 2521 CGCAACGGGAGATCTCTTACACCATCCGCAAGTGGGACTTATGCAAGGATTTATC 2580

QY 1467 Arg-AlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgAlaAla 1486

Db 2581 AGATGCG----- 2587

QY 1486 aValValIleGlnArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyr 1506

Db 2587 ----- 2587

QY 1506 rGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526

Db 2587 ----- 2587

QY 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnIly 1546

Db 2587 ----- 2587

QY 1546 sPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSe 1566

Db 2587 ----- 2587

QY 1566 rMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLe 1586

Db 2587 ----- 2587

QY 1586 uMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnAr 1606

Db 2587 ----- 2587

QY 1606 gTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeuAspIleAr 1626

Db 2587 ----- 2587

QY 1626 gLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLysVa 1646

Db 2587 ----- 2587

QY 1646 lGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaValIle 1666

Db 2588 -----TTAGCAGTGGT 2598

QY 1666 uSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPh 1686

Db 2599 AAGTCGATTGGATCGTCTTTTCGGCAGCTGTGCACACCTGCTCATGTGGTGTTCGGAGTT 2658

QY 1686 eMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGI 1706

Db 2659 CATGTCACCGTTTGTCTATGGCATCATGGCTCAGGCCATTGCATCAGAGGTGGAAGCA 2718

QY 1706 nLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrVa 1726

Db 2719 GCTTATAGAGCGCTGCAGCGGATCATCTTAATTTGGCCCGCTACAAATGACACCGGT 2778

QY 1726 lAsnThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAs 1746

Db 2779 GAACACGTTCCAGAGGCGGTTTGGTCCCAITGGCCAGATGTTATTGGCTGCTGCGA 2838

QY 1746 pLysAspSerGlnIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysPr 1766

Db 2839 CAAAGACAGTGAATATTCACACCTTTGTGCACCCCTCAATTTGGGTATTCGCCCACTGTCC 2898

QY 1766 oLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIleTyrMetValAr 1786

Db 2899 TAAAGCGAAGATCATTCACGACTATATGACCAACCCAGAGGCCATTTACATGGTGGC 2958

QY 1786 gGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaArgLysProPr 1806

Db 2959 CGAAACTAAGAGCTCGTGGCCCGCAAGGAAAAGATGAAGCAAAATGCACGCAAGCGCC 3018

QY 1806 oProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrProCysSerLeuPr 1826

Db 3019 GCCAATGACAAGTGAAGCGGTATGAAGAGCCAAAGATAAACTTCACGCGGTTCCTGTCC 3078

QY 1826 oSerLeuGluProAspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerVa 1846

Db 3079 CAGCTGGAGCGGACTTCGGATCATCGCTACAGTCCCTACAGTTTATCTGCTCCGT 3138

QY 1846 lTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMetPhe 1861

Db 3139 TTACGCTTCGATACGATTTTGTGCAAGCTGCAGATCGACATGTTT 3184

RESULT 6

AF533752 9846 bp mRNA linear ROD 17-OCT-2002

LOCUS Mus musculus abnormal spindle (Asp) mRNA, complete cds.

DEFINITION AF533752

ACCESSION AF533752

VERSION AF533752.1 GI:24079963

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 9846)

AUTHORS Bond, J., Roberts, E., Mochida, G.H., Hampshire, D.J., Scott, S., Askham, J.M., Springell, K., Mahadevan, M., Crow, Y.J., Markham, A.F., Walsh, C.A. and Woods, C.G.

TITLE ASPM is a major determinant of cerebral cortical size

JOURNAL Nat. Genet. 32 (2), 316-320 (2002)

MEDLINE 2242604

PUBMED 12355089

REFERENCE 2 (bases 1 to 9846)

AUTHORS Mochida, G.H. and Walsh, C.A.

TITLE Direct Submission

JOURNAL Submitted (31-JUL-2002) Neurology, Beth Israel Deaconess Medical Center, 77 Avenue Louis Pasteur, HIM 807, Boston, MA 02115, USA

FEATURES

source 1..9846

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="Swiss Webster"

/db\_xref="taxon:10090"

1..9846

/gene="Asp"

138..9506

/gene="Asp"

/codon\_start=1

/product="abnormal spindle"

gene

CDS

```

/protein_id="AAN46088.1"
/db_xref="GI:24079964"
/translation="MATHQASCPBERGRARRPDPPEAGDPSPPVILLISHFCGVPFLCF
GVRVGRTRSLVNLHPHEEPLOVELSLRAAGQGFVAPNRCCLPKPKKILITISVTW
TPLEGGVRIIVTFLVNDFLKHQAILLGNABEPPKKKSLWNTSKKIPASSKHSTCLP
KNQHNESFTISOKDRILRSPLQPCENLAMESECSPTEKNKVPISIRRECSQSETCPL
LFTRESTAYSLSHSENTONLKVODASISOTFDNEBVANETFINPLSVCHQSGEDRK
LTLAPSCSSPLNSTQTQIHLSPLDSFVNNRYTSDNDLKSMMKNVLSDFPRKDPAESVCL
ESQTHVEVCQTILSPFLNDNGLKGLKGLNPKSVNPLSTPQFVKDSMGHVQQTGKS
NEASQDMINEGLAFTPECHOAQTSPSRSEKQNPVEVKPHKIDYTKQPKCEFDQAF
CHOSQPKHPRILSATVTRKPTNAREKLPEINPKDAKRCLEGLVQORKEVGSLEAF
KRIHREPLSTSTVARTTKKGLTKSKTSSILSRGLSKKKMDSILKTPLSKTKKCR
RSIVAAQSHLTFIKPLKAALPHRPMPPAAKNMFDYDERWEKEKQOGFTWMLNYLTDP
DFTVKTNVSKKVAEIEVLGRVLKRDHLKQIQGQKQVNLWLLSINPLRLTGLSET
FTSEKMKVAKVLEIEVLGRVLKRDHLKQIQGQKQVNLWLLSINPLRLTGLSET
VFGLILCLFNDHAKTSLRIDPDCLFCCKDAEFKASKELLAFPSRFLDSGEGDLSRHLF
KLILLICLFDHAKTSLRIDPDCLFCCKDAEFKASKELLAFPSRFLDSGEGDLSRHLF
LGLPVSHTQPLDDEFAVNLAVDLQCGVRLVTELLTONNLSLKLIPALSRVQ
KMNVDLAVOLVRSRGLPDTDEHGSATSSKDVDRHREKTLGLAKIALAFODVISLN
IDQKESIDFLKTHHSIKRAMSALTCPSQALTNQDKRIISGNFERVGSVOLLMWV
NAVCAFYNNKGVNFTVSDGRVLCYLIIHYHPCVFPDAICQRTSOSVACAQGSVV
LNSSESGEGCLDLSALDEHSTPEMYKELLEKNFHLVRSAAADLPAMIH
SDMNTITPDEKVTIYLSFLCARLLDLKEIRAAALITQTTWRKYLKRLDKLHQERDK
AARYIQSVLNFISRRLOKNVSAALVTQKWRVRSQKRLMKLNEKLAKLQNSAV
LLOAWRYSFKREFLRLKHYSVILQSIIRMKIALTSYKRYLWATVTIORHWAYLSG
KRDQOIFRKLSSSILVITQPMFRWRKRLQLOTKAANTLOPAPREHLRKQIRERSAV
VIQSWYNRHRELQIYIIRSCVIVIQRRVRCFOAKLYKRRKDALILQKHYRQKGS
KLHADYLOKAAATLOQAFRGMKARHSYRLOIGAACVLSQSYWRMQRVRFNLKK
MVILOAHIRYQLOQYKTKKAAITIQTHFRASISARRVLASQYKTRSSVILQSA
CRGMAKAPFHALASVIKIOSYVYVIRKTFONFNKNTKLQSIIVMKQSRQYLO
QSYFMETARQYLKCKAALVIOFYCAVEPAQIOSKNFLQVKEALICLOAAVGGCK
VRRLQKQSTAAVTIIOVRFGHSQRMKTQMLQSAVKLOPWRKQVAYDNRIOFLKT
RAVALFORWRSOKLASOKREYIOVRESKIKLOSHFRGLVRKOLRLOCKAAISL
RAEVCVLSQSVARQVQOLRQHEAAVKIOSTFRMAVAQOQYKLLRAAAVITQOHRV
ARAAGTKLAIYIOLRAALVFOQAAGKMLRQIARVHOCAALIOSYRMYIQRKXW
SIMTKAALQIOLCYRQVYKQERHLYLKAAVITQOQOYLNLRRAVVOVQAAYRG
TIQSKFRAYRTQKVTYTSATVIOVRWYNIKITQOQOYLNLRRAVVOVQAAYRG
IVRPRRTOHMHMATHLEAMFMRQSRVRLVKMTAALIOVRYRAYVLGKIQHEKYL
RTLKAIKLOAGVGARVTRVNMKHAATLLOSHFRGHROQTYFHLRKAATWVQR
YRAVKEGSAEIVQSRSLRSVLLIOAPFRGLTRRLHKLAMHLAATLQRRPRTFAMR
KFLSLRKTAIMIQOYRQYRQYKYSROQLLEKAVIKIOSYRGMVVRKVRQKHRAA
VIOATFMHGAQVRYOHLKRAVYIOVHTAAELQORKAAVILQAAVGMKTRSHLK
TMHSSATLQSOFAFIVRRRPIALRKAALFOVKFRATLVAKHKLHOFLOKRAAIT
IOSYRMLVWQKIQEMHRAALLOAFEMHRTVVAFHWKASIRIOCCYRTVETIK
LOKELKLEBQSHSAVILQSTYRMTQRCFTQORWAKVITQKTYRANKRQDILLYCA
KEETPLLOMHQGLNTAKQROHGAAMITQKHFRAFKAARLMEARFGQACRKYK
KKYLSKVBAACRIQAWYRMRRAHKYLLTLLKAVNIEGYLSAQALARRPFLKORAAAI
IORKWARTLSVRGARENLRHRAACVIOAHFRGQARQSFLOQRSANVILIOHVRWAV
AAKQRIKVIKLSKSTVVVOALVRGWLVRKRVSEOKATRLPFPHTAAVCHMCKALQI
RAYRLHVTLRNAKHMDSVIFIORWFRLOKRFLEQYHKLISPTREAHACWLQODR
AASVLOKARFPLCRREKITSNATRIQALWRGYSWRKKNNDHEIKALRSRAVST
TVEENKLYRTERALHLLTYKHLSDALDKLEVVTRLSLSPGCCENNAESGAVSTI
FVATRSNRPVCEMVGVAVQVLINAKYDKTAAVYEAENCVDITLLELLQVYREKP
GDRVAEKASIFTRCTCLLAVLLKTEQCAFQASQSKVTDRIYRLKFTVPKHKVNTQ
GLEFDKQKNSCVGPECIPIERTMKTRLVRLKQVWVLRDRDNVEEITNSIQALQVLVMDTL
GISI"

```

ORIGIN

Alignment Scores:

Pred. No.:	2,4e-67	Length:	9846
Score:	1374.00	Matches:	552
Percent Similarity:	38.11%	Conservative:	371
Best Local Similarity:	22.7%	Mismatches:	741
Query Match:	14.4%	Indels:	758
DB:	10	Gaps:	77

US-09-914-698-1 (1-1861) x AF533752 (1-9846)

QY	1	MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeu	20
Db	429	ATTTCCGTTACCTGGAGCCACTGCGAGAAAGGGGAGTGGGGAGATTGTCACATTTCTT	488

QY	21	AspAsnArgAsnPheArgLysGluValMetIleLeuLeuLysSerLysSerAsnGlnPro	40
Db	489	GTRAAAT--GATTTCTGGAAGCACCAGGCTATATTACTA--GGAAATGCGAAGAGCCT	542
QY	41	ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr	60
Db	543	AAGAAGAAAAAGAGAACTCTTTGGAATACCAAGTAA--AGATTTCCAGGCC	590
QY	61	GlyAlaGlyLysThrMetLysSerVal--ValSerAlaAlaVal	74
Db	591	TCCTCAAAACATCAAAAAGGACTTCCAAAACCAACATTTTAATGATCATTTACTATT	650
QY	75	GlnGlnLysLysArg--MetSerAla	82
Db	651	TCACAAAAACAGACAGATTAGGAGCCACTGCGAGCTTGTGAAAAATCTGGCTATGAGTAA	710
QY	83	AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSer--ArgProAla	101
Db	711	TGCTCTTTCCCAACAGAAAAAC--AAAGTCCCCACCCCATCCATTTAGTCTTATT	761
QY	102	AlaTrpAlaHisProProGlnAlaProLeuVal--GluLysAsnValTrpLys	119
Db	762	AGAGAAATGCCAGAGTGAACCTTGCTGCCACTGTTTTCGCGAGTCCACTGCCTATTCA	821
QY	120	ThrProGlnGluProValTrpLysSerProGlnProArgSerLeuLys--GluAsn	138
Db	822	TCTCTTCATGAA--TCTGAAAAATACACAAAATTTAAAAAGTACAAGAT	866
QY	139	LeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuAspPhe--Thr	157
Db	867	GCCAGCATTTCCAAAACTTTTGATTTAATGAGGAAGTCGCAATAGAACTTTTATTAT	926
QY	158	ProLeuThr--GluThrArgGlyLysGlyGlnAlaThrIlePheProAsnAsn	174
Db	927	CCCATAGTCTGTCTCACCAGAGTGAAGGATAGGAACTCAGCTTGGCCCCAAAC--	983
QY	175	LeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAsp--	193
Db	984	--TGTTCTTCCCTTTGAATAGTACACACACCCAAATACACTTTTCTAAGTCCAGATCT	1040
QY	194	--MetArgProArgGlyIleThrProAspAspLeuGluAsp--GlnProAlaThrAsn	211
Db	1041	TTTGTAACACACAGGTATACGCTCTGATATGACCTGGAATCAATGAAATAATGTTTATCA	1100
QY	212	LysThrPheAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp--	230
Db	1101	GACACATTTAGGAAGATCCCGCAGAGTCTGTATCTCTGGAATCCCAAACTGTACATGAA	1160
QY	231	--CysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal	249
Db	1161	GTTTGCACAAACAATC--TTAAGCCAGATCTTTTAAAAATGATAATATGAGACTTAAA	1217
QY	250	HisAlaThrHisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProPro	269
Db	1218	AAAGGCTAAACTTCAATCCGTC--AATCCCGTT	1250
QY	270	ArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySer	289
Db	1251	TTTGCGCCACACAAATTT--GTAAAGGATAGCATGGGACACGCTAGCTCAG	1298
QY	290	ProLeuArgLysTyrSerGluSerMetLysAspLeuSerLeu--Leu	304
Db	1299	CAAAAGGGTAAAGTCAATGAAGCTTCTCAGGATTTGGAGATAAATGAAGTTTGGCATAT	1358
QY	305	SerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIle	324
Db	1359	ACCCCTGAGTGCAGCATGCA--CNAATCCAGCAGCAGATCTGAAAAACAA	1409
QY	325	ArgSerIleGlu--	328
Db	1410	RACCCCTGTAGAGTGAAGCCACATAGTATGATTTTACAAAGCAAAAGCCCAAGACTTGT	1469
QY	329	-----GlnAsnArgTyrTyrGlnGlnGlnGlnIleGlnIleLysAlaLysAspLeuAsn	346



Db 1470 GAGTTTCAGGATGCTTTTTCACACGAGTAACACACCGCACAGACGCCCAATTCCT 1529  
QY SerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeu-----PheAsn 364  
Db 1530 TCGCCACCTTTACGAAAGAAACACACCAATGCGACAGAAAGTACCTGAGATTAAAC 1589  
QY HisSerGlu-----IleLeuAlaGlnSerSerSerArgPheAsnLeu 377  
Db 1590 AACCCAGACCCAAAGAGATGCTCGAAGGTCTAGTAGTCAACGTGGAAAA----- 1640  
QY HisGluValGlyArgLysSerValLysGly-----SerProValLysAsnPro 393  
Db 1641 ---GAAAGTGGGTAGCCTGAGAGAGAAAGGTTTTCATCCTCTCTCTCCCGTTGCGAGCCA 1697  
QY 393 ----- 393  
Db 1698 GGAGTGAGTAAGCTCTGAGCTACAGGATGAAGTAACGCGCTGCACTGTTGCTTGCCT 1757  
QY HisLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyr 413  
Db 1758 CGGAAAGAAAGAGCCACGGA---ACAGTGGGAGATGCA-----AATGAAAGATGACA 1808  
QY 414 ArgAsnGluThrValAlaIleSerProProLysLys-----GlnArgValGluAsp 430  
Db 1809 GCTGAAGATGGATGGACATGTGTGAAGTCAAACGAATCCATTTTCTCTCTGGAATCT 1868  
QY 431 Thr-----ThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSer 448  
Db 1869 ACACCTCACTGTGTCAGAACACCAAAAGAAAGAGGGCTTACCTCAAAGCGGATAAGC 1928  
QY 449 AlaHisAlaTribProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeu 468  
Db 1929 AGC-----CTGGAGCGGTGAGGCTGAGAGAGAAATGGAATTCATCAATA 1973  
QY 469 MetLysLysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLys 488  
Db 1974 TTAATAAACACCTCTTCTAAGACAAAGAAAGAGAGATCCATT-----GTTGCT 2024  
QY 489 LeuTyrAspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThr 508  
Db 2025 GTAGCACAGTCTCATTTG-----ACCTTTATAAAGCCATTAAGAGCAGCTATTCCT 2075  
QY 509 ThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValaspArg 528  
Db 2076 AGACACCCATGCGTGTCTGCAAGAACATGTTCTATGACGACGCTGGAAGGAAG 2135  
QY 529 HisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAsp 548  
Db 2136 CAGGAACAGGGCTTCACTTGGTGGCTTAATATATATTAACA---CCTGATGACTTCACT 2192  
QY 549 AlaAspLeuAsn---AsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLys 567  
Db 2193 GTAAAAACCATGTTTCTAAAGTGAACGCTGCGAGCTTTGTTTGGAGCGGAAAGCCAA 2252  
QY 568 GluLeuVal-----ValAlaProThrLysGluGluGlnSerMetAsn---TyrLeu 583  
Db 2253 CATAAAAATAGTGTCCGGAAGACCTACTAAGAGAGAGTGTCTCTCAGAGCTTATACA 2312  
QY 584 ThrLysTyrArgGluThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGln 603  
Db 2313 GCCAGCTGCGGCTGAATAGATTACGCCGTACAGCTTGTCTCTGTTTACTTCTGAATAA 2372  
QY 604 MetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArg 623  
Db 2373 ATGGTTAAGCTATTAAAAAGGTGAAATCGAAATTTGAAGTTGGGCGGCTACTTGTGCA 2432  
QY 624 SerAspArgAsnLeuHisLeuAspValMetGlnArgThrIleLeuGluLeuLeuLeu 643  
Db 2433 AAAGACCGACACTCTGGAAGAGATATTGACACAGCCAAAAGTCTCTGAATGGGTCTTGG 2492  
QY 644 CysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysIleGln 663

Db 2493 TCATATAATCCTCTCTGGCTTCGAATGGTCTGGAGACAGATTTTGGAGAACTGATCCCT 2552  
QY 664 MetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArg 683  
Db 2553 CTGGCAGACAAAGTACAGTACGGGATTCAGGATGTTTCAATCTGTAATCGCTTACTTTGG 2612  
QY 684 AsnLys-----CysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu--- 700  
Db 2613 AATCTGACATAGCAGCTGAATATATAGACCCCACTGTTCTCTCTCTCTGCGGATGGT 2672  
QY 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuLeuProPheLeu 720  
Db 2673 CATGAAGCAGCTTGTCTAAGTTCACATTGAAAAAATGCTGTGTGTGATCTGTCTTCTT 2732  
QY 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
Db 2733 GATCATGCTAAAAATTTCCAGACTATTGATCATGATCTTGTCTCTCTCTGTAAGATGCT 2792  
QY 741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIle 760  
Db 2793 GAATTCAGGCTAGCAAGAGACTTCTTAGCCTTTTTCAGGGATTTCTCAGCTGGTGAA 2852  
QY 761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
Db 2853 GGTGACCTTCTCGTCACCTTAGCTTCTTGGGTTTACCTGTGTGAGCCACGTTTCAGACCCA 2912  
QY 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValaspLeuArgAspGlyValArg 800  
Db 2913 CTTGATGAATTTGATTTTGTGTACAAATCTTGCAGTGGACTTGCAGTGGGTTCCG 2972  
QY 801 LeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
Db 2973 CTTGTCCGCACTGTGAGCTTCTCACACAGAACTGGAATCTGTCCGACAACTCAGGATC 3032  
QY 821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
Db 3033 CTTGCAATAGTTCGAGTTCAAAGATGCAATGTTGATCTCTGCTCTCAGGTTCTCAAG 3092  
QY 841 GluAlaAsnPheGlnLeu-----GlyGlyAspIleAlaAlaGlnAspIleVal 856  
Db 3093 TCTCGGGAGTTCACACTGACTGACGAGCATGGGCGCTATCTCGCTAAGGATGTTGTG 3152  
QY 857 AspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeuIleTyrLysPheArg--- 875  
Db 3153 GACAGCACAGAGAGAAACCCCTTGGTGTCTTGGAAAAATAGCATTTGGCTTTCCAGGTG 3212  
QY 875 ----- 875  
Db 3213 GATATTTCCCTTAATTAGATCAATTAAGAAGAAATTTGATTTTCTAAAGCACACAC 3272  
QY 875 ----- 875  
Db 3273 AGTATAAAGAGAGCTATGTCTGCACTAACATGCCCTTTCACAAGCTATTACTAATAAGCAA 3332  
QY 875 ----- 875  
Db 3333 AGAGACAAAAGGATTAGTGGGAACTTTGAAACGCTATGTTGACAGTGTGAGTGTGATG 3392  
QY 875 ----- 875  
Db 3393 GATTGGGTAAATGCAAGTTGTGCTCTTATATAAAGGTGGAGAAATTTTACAGTGTCT 3452  
QY 875 ----- 875  
Db 3453 TTCTCAGATGGCGGCTACTATGCTATTGATPCCACCATTAACCCCGTGTATGTGCT 3512  
QY 875 ----- 875  
Db 3513 TTTCGATGTATATGCGGACGACCACTCAGTCCGTGGCTCGGCACAGACTGGTTCAGTG 3572  
QY 875 ----- 875  
Db 3573 GTATTGAATTTCTTCATCGGAATCTGAAGAGGTTGCTGTGGATCTGTCTCTTGAAGCGCTT 3632

QY 876 -----SerProLysPhe 879  
Db 3633 GATCAGAAAGTACTCCAGAAATGTACAAAGAGCTTCTAGAAAACGAGAAAGAAATTTTC 3692  
QY 880 His----- 880  
Db 3693 CATTTGGTGGATCTGCAGCTAGGATCTGGAGGATACCTGCTATGATTTCATCAATCA 3752  
QY 880 ----- 880  
Db 3753 GACATGTCAAATACAATTCAGAGAAAAGGTGTTATCACCTATTTGTCTATTCTTTGT 3812  
QY 881 -----AlaAlaAlaThrValLeuGlnLysTyrTyr 890  
Db 881 ----- 881  
Db 3813 GCAAGGCTTTTGGATCTCCGTAAGAAATAGAGCTGCTGCATTTATACAGACAAATGG 3872  
QY 891 ArgArgHisTyrLeuHisValValIleGlnArgIleArgHis----- 905  
Db 3873 AGAAATATAACTA-----AAAAGATCTGAAACACCATCAGGAAAGAGAC 3920  
QY 906 -----LysGluLeuMet 909  
Db 3921 AAAGCTGCTAGAGTTATTAGTCAGTGTGTTCTCAATTTTCTAAGTAGACGAAGATTGCAA 3980  
QY 910 ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys 929  
Db 3981 AAAAAGCTCAGTCAGCTCTGCTCATTCAGAAATGCTGGCAGAGCTCTAGCCCAAGG 4040  
QY 930 TyrValLysLeuPheLysThrGluArg-----ThrGlnAlaAlaIleIle 944  
Db 4041 AAGTTGGGATGCTGAAAATGAAAACCTGCGCAAACTTCAAAATAAATCAGCAGTGTCTC 4100  
QY 945 LeuGlnLysPheThrArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHis--- 963  
Db 4101 ATTCAGGCTTATTGGAGAGTACTCCACCAGAAAACGGTTTCTCAGGTTGAAGCACTAT 4160  
QY 963 ----- 963  
Db 4161 TCAGTCATCTCAGTCAGTAAAGATGAATGCTCTTACATCTTCAAAAGCATAT 4220  
QY 964 -----SerIleIleThrIleGlnArgTyrPheArgAlaGlnLeuGlyArgGlnHis 981  
Db 4221 CTTTGGGCCACAGTTACATTTACAGGCGATTTGGCTGCTATTTAAAGTGGAAACAGAT 4280  
QY 982 ArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGln-----ArgIle 998  
Db 4281 CAGCAATATTTTAGGAGCTGAATCTCTCATCTCTGTAATCCAGTTTATGTTTCCAGACGG 4340  
QY 999 TrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGln 1018  
Db 4341 TGGAGCGACGCAAGCTCAGTGTACAAACCAAGCTGCAGTAAACATTCACGCGAGCTTTT 4400  
QY 1019 ArgSer-----GlnLysGln----- 1023  
Db 4401 AGAGATGGCATCTCAGGAAACAAATACGAGAAAGATCCGCTGTGTGATACAGTCAATGG 4460  
QY 1023 ----- 1023  
Db 4461 TACAGATGCACAGGGAATTCAGAAATACATTTATATTAGCTCTTGTGTTATTGTCTATC 4520  
QY 1024 -----GlnAlaAlaSerTyr----- 1029  
Db 4521 CAGAGAGAGTTCCGTGCTTTCAGCCCAAGTTATATAAGAAAGGAAAGATGCCATC 4580  
QY 1030 -----IleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPhe 1047  
Db 4581 CTGACTCTCCAGAAACATTTACAGGCGCGCTCAGAAAGGGAAGCTACGACACGCTGACTAT 4640  
QY 1048 LeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArg----- 1062  
Db 4641 TTGCAGAAACGAGCTGCCACCATTCGACTGTCAGGCTGCATTCAGGGGATGAAGGCTCGC 4700

QY 1063 -----SerLysTyrSerMetLeu 1068  
Db 4701 CATTCATACAGGTTGCAGATTGCAGCTGCCTGTGTCTTCTCAGTCTCTATTGGAGATGAGA 4760  
QY 1069 GluGlnArg----- 1071  
Db 4761 CAGGAAGAGTTAGATTTTAAACCTTTAAAGATGGTTATTAAAGTTGCAAGCACACATA 4820  
QY 1072 LysGluPheGlnGlnLeu-----LysArgAlaAlaIleAsnIleGln 1085  
Db 4821 AGAATATACCAACAACATACAGAATATAGAAGATAAAGAACGACCATTTACATTTAG 4880  
QY 1086 GlnArgTyrPheArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArg 1105  
Db 4881 ACTCATTTCCGAGCTTCCATTTCCAGCCAGGAGGTTCTAGCATCTTTATCAGAAACCAAG 4940  
QY 1106 SerSerValLeuLysValGln-----AlaTyrArg 1115  
Db 4941 TCTTCAGTCATTTGTTCTACAGCTCTGCATGTAGAGGATGCAAGCAGAAAGCAATTTCCG 5000  
QY 1116 LysAlaThrIleGlnMetArgIleAspArgAsnHisTyr-----TyrSerLeuArgLys 1133  
Db 5001 CATGCTTTAGCATCCGTTATAAGATTCAGTCATATTATCGGCTTATATATGTAGGAAG 5060  
QY 1134 -----AsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGlu 1150  
Db 5061 ACATTTCAGAACTTCAAAAATGCCACAATAAAGCTGCAGTCAATTTGTCAAAATGAAACAA 5120  
QY 1151 GlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArg 1170  
Db 5121 TCACGGAAGCAGTATTTACAAATAAGAGCAGCTGCACTGTTTATCCAGCGTGTACCGT 5180  
QY 1171 MetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle 1190  
Db 5181 TCCCAAAACCTGGCTCACAAGAGAGAAAGAGTACATACAGGTGCGTGAATCTGTATC 5240  
QY 1191 AsnValGlnArgArgTyrPheArgAlaThrLeu----- 1200  
Db 5241 AAATGTCAGTCACACTTTAGAGGATGCTGTCAGAAAGCAGTGTGCAATGTAAATG 5300  
QY 1201 -----GlnMetArgArgGluArgLysAsnTyrLeuHis 1211  
Db 5301 GCTGCCATTTCTPACAGTCTTATTTCAGAAATGAGAACAGCGAGCGCGCTACTCTGAAG 5360  
QY 1212 LeuGlnThrThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLys 1231  
Db 5361 ATGTGCAAGCGGCTCTTGTCATTCAGAGTTTCTACTGTGCTACAGAGCACAAATCAGT 5420  
QY 1232 GlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValValGln----- 1247  
Db 5421 CAAAGGAAGAACTTCTTTCAGGTCAAAAGAGCAGCTATTTGTTGCGAGCGAGCTTACAGA 5480  
QY 1248 -----LysArgArgArgAlaLeuLeuGlnMet----- 1256  
Db 5481 GGCTGCAAGTACCGCGCAATCAAGCAGCAGTCTACGCTGCTGTGTACATTCAGAGG 5540  
QY 1257 -----ArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrIleLys 1272  
Db 5541 GTTTTTCAGGCCACAGTCAGAGGATGAATACACAGACTATGCTTCAGTCTGCTGTGAAG 5600  
QY 1273 LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly 1292  
Db 5601 ATTTCAGAGATGTTGAGGCGCACAGAGGTTTCTTATGATGAGAAATTTCAATTTCTAAAG 5660  
QY 1293 ThrGlnAlaAlaValSerCysLeuGlnMetHisTyrPheArgAsn----- 1306  
Db 5661 ACTAGAGGCTGTGTTTGGCTTCAGTCTGCTTACCGTGGCTGCAGGTTTCGGCACACAG 5720  
QY 1306 ----- 1306  
Db 5721 CTCAGAGGCAACATGAGGCTGCAGTGAAGATCCAGTCTACGTTTGAAGATGGCCGTGGCC 5780  
QY 1307 -----HisLeuLeuArg----- 1310





```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
1..10434
/notes="ASP"
/codon_start=1
/product="abnormal spindles"
/protein_id="AA040011.1"
/db_xref="GI:23884551"
/translation="MANRVRGRCWGVSPTRRRPAGLRGPAEEERASSPPVLSLSPH

```

CDS

```

Alignment Scores:
Pred. No.:      1.6e-66      Length:      10434
Score:          1360.00      Matches:    478
Percent Similarity: 39.81%    Conservative: 362
Best Local Similarity: 22.65% Mismatches:    630
Query Match:      14.29%     Indels:      640
DB:              9          Gaps:        67

```

US-09-914-698-1 (1-1861) x AF509326 (1-10434)

```

Qy 5 TrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuIleAspAsnArgAsn 24
Db 328 TGGACACCATCTCAAGAGAGCCGAGTAGAGAGATTATGACATTTCTGTAAAT--GAT 384
Qy 25 PheArgLysGluValMetIleLeuLysSerLysSerAsnGlnProValLysAsnPro 44
Db 385 GTTCTGAAACACCAAGCTATATTACTAGGAATATGCAAGAGCAGAAAGAAAAGAGG 444
Qy 45 ArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThrGlyAlaGlyLys 64
Db 445 AGTCTTTGGGATACCAT-- 462
Qy 65 ThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMetSer-- 81
Db 463 ---AAAAGAGAGAAATTTTCCAGCTCTACAGTCACAAACAGAGGGTTTCAATATTCAG 519
Qy 82 -----AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArg 99
Db 520 AATGTTAATAAAACATTTAGTGTGTTTCCAAAAAGTTGACAGAGTTAGAGCCCACTCAA 579
Qy 100 ProAlaAlaTrpAlaHisProProGlnAlaProLeuValGlnLysAsnValTyrLys 119
Db 580 GCTTGTGAAACTTGGCTATGTAATGAAGGCGGTCTCCCAACAGAAACATATCTTTA--- 636
Qy 120 ThrProGlnGluProValTyrIleSerPro----- 130
Db 637 ATACTTGAAGAAATAAAATACCCATCATCTATAGCCCTGCTTCAATGAATGCCAT 696
Qy 131 -----GlnProArgSerLeuLys-----GluAsnLeuSerPro 141
Db 697 GGTGCAACTGCTTGGCACCTCTCTGTACGTCGATCTACTACTCATCTCTCTTCAATGCA 756
Qy 142 MetThrProGlyAsnLeuLeuAspVal-----IleAspAsnLeuArgPheThr 157
Db 757 TCAGAAATAAGGAACTATTAAATGTATACAGTGCACACGTTTCAAAAGTTTCTTTAAT 816
Qy 158 -----ProLeuThrGluThr-----ArgGlyLys 165
Db 817 GAGAAAGCTGTAACCTGAACTTCCCTTAAATCTGTAATGTTAATGCGCAAGAGGAGAG 876
Qy 166 GlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThrProLeuLysGly 185
Db 877 AATAGTAACCTT-----AGTCTTACCCCC 900
Qy 186 AsnValLysSerCysAlaAsn-----AspMetArgProArgArgIleThrProAspAsp 203
Db 901 AACTGTTCTTCAACTTTCAGACATTACACAAAGCCAAATACATTTTCTTAAGTCCAGATTCT 960
Qy 204 LeuGluAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsn 223
Db 961 TTT-----GTAATAATATGTCATGGAGCTAAT 987
Qy 224 IleSerLeuAspThrLeuAspCysSerArgIleAsp-----GlyGlnProHisThr 240
Db 988 AATGAACCTAGAAATTAGTAACATGCTTTTCATCAGATATGTTATGAAAGATAATTACAG 1047
Qy 241 ProLeuAsnLysThrThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle 260
Db 1048 CCTGTGTCATTCGAATCAACAATTCACATGAAATTTATCGAAATTTTA----- 1098
Qy 261 HisGluGluGluGlyProSerProArgThrProThrLysSerAlaIleHisAsp--- 279
Db 1099 -----AGTCCA-----GATTCTTTCATAAAGATAAT 1125

```

ORIGIN

Qy	280	-----LeuLysArgAspIleLysLeuValGly---SerProLeuArgLysTyrSerGlu	296
Db	1126	TATGGACTAAATACAGGATCTAGAATCAGAGTCAGTTAATCTATCTATTTATCCCTTAATCAA	1185
Qy	297	SerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMet	316
Db	1186	TTTTTAAAGATAACATGCGCATATATGTGTACATCTCAGCAAAACATGTAAAGTACCATA	1245
Qy	317	ProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGluGln	336
Db	1246	TCAAATGAAAATTCCTCAAGTCCCAAGTCTCCTGAAGAT-----TGGAGAAAAAGT	1296
Qy	337	GlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAla	356
Db	1297	GAAGTTTGGCCACGATATTCTCGAATGTGAGGGTTCAAAATCTCCCAAGCTATTTTGAA	1356
Qy	357	GlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSer---ArgPhe	375
Db	1357	GAAC TAGTAGAANAAGTAGTCAAATTCACAGTTTTATAAACAATAATCCTTAATTT	1416
Qy	376	Asn---LeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnPro---	393
Db	1417	TCTGCAGTTCAGGATATTCTAGTCATAGCCACATAAACCACTTAAGAGACGTCCTAATA	1476
Qy	394	-----HisLysArgArgSer-----	398
Db	1477	CTTTCTGCCACTGTTACTTAAAGAGAGGCCACCTGTACCAGAGAAAAACCAACTGAGATT	1536
Qy	399	-----	403
Db	1537	AATAAACCAAAAGCAAAAGATGTCACAGTCGATGGGTGACATGAAAGATTAATA	1596
Qy	404	SerAspAlaProSerAsnGluSerLeu-----	412
Db	1597	AATAATCAAAGGAAAAGAGATTTTCATTCTTATCTTCCAATTATAGATCCATATTA	1656
Qy	413	-----TyrArgAsnGlu-----ThrValAlaIleSer	421
Db	1657	AGTAAATCTAAGAGTTTATAAAACGAGGTAAACACCCCTCTTCGACACAGCTTCAGTTGCT	1716
Qy	422	ProProLysLysGln-----ArgValGluAspThrThrLeuProArgSerAlaAlaPro	439
Db	1717	CGGAAAGAAGACCGATGGAGCATGGAGATGGAAGATGCAAAATGTG---AGAGTTGCAATTACA	1773
Qy	440	AlaAsnAlaSerAlaArgSer---SerSerAlaHisAlaIleProHisAlaGln-----	456
Db	1774	GAACATACAGAAGTGCAGAGAAATCAAAGAATCCATTTTCTCCCTCAGAGCCTAAACA	1833
Qy	457	-----SerLysLysPheLysLeuAla	463
Db	1834	TCAGCTGTTAAGAAAAACAAAAATGTGACAAACACCCATCTCAAACGTAATTAGCAACAGA	1893
Qy	464	GlnThrMetSerLeuMetLysLysPro-----AlaThrProArgLysVal	478
Db	1894	GAGAAATTAACCTTGAAGAGAAACATGATTATCAATATTTCAGACTCCAAATTTCTPAAA	1953
Qy	479	ArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyr-----	495
Db	1954	ACAAACAAAAGGACAAAACCATATTCGCTGTGGCACAGTCCAGTTTGACCTTCATAAAA	2013
Qy	496	---MetGlnThrCysIleAsnProAspProPheAlaAlaThrThrThrIleAspProPhe	514
Db	2014	CCATTAAAAACAGATATTCCAGACACCCGATG-----CCATTTT	2052
Qy	515	LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLys	534
Db	2053	GCTGCAAAAAACATGTTTATGATGAACGCTGGGAAGGAAAAAGCAGGACAGGCGTTCAC	2112
Qy	535	LysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspPheAsn---	553
Db	2113	TGGTGGTTAAATTTTATATTAAACC---CCTGATGACTTCACTGTAAAAAACAAATATTTCT	2169

QY	554	LysLeuAspValGluLysLysLeuPheAsnGluValValArgAsnLysGluLeuValVal	----- 571
DB	2170	GAAGTAAATGCTGCTACTCTCTTTTGGGAAATAGAGAATCAACATAAAATAGTGTTCCT	2229
QY	572	--AlaProThrLysGluGluGlnSerMetAsn--TyrLeuThrLysTyrArgLeuGlu	589
DB	2230	AGAGCACTACAAAGAGGAAATGCTCTCAGAGCTTATATCTGCTCCGGTGTAGGTAAAC	2289
QY	590	ThrLeuArgLysAlaLaValGluLeuPhePheSerGluGlnMetArgLeuProCysSer	609
DB	2290	AGACTAGCTCGTCGACGATCCGGTTTCTTACCTTCGAAAAATGGTTAAAGCTATTAAA	2349
QY	610	LysValLaValTyrValAsnLysGlnLaLeuArgLileArgSerAspArgAsnLeuHis	629
DB	2350	AAGCTCGAAATTTGAATTTGAAGCTAGGCGGTTTAAATGTTCCGAAAAATAGACACCTATGG	2409
QY	630	LeuAspValValMetGlnArgThrLileLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	649
DB	2410	AAAGATGTGGGAGAGACCTCAGAAAGTCTGAAATGGCTGTGTGCTCTACAAATCTTTGTGG	2469
QY	650	LeuArgLeuGlyLeuGluValValPheGlyGlyLysIleGlnMetGlnSerAsnArgAsp	669
DB	2470	CTTCGAATTTGCTAGAGACCACTATATGAGAGAACTCATATCTTTGGAGATACACGTGAT	2529
QY	670	IleValGlyLeuSerThrPheLileLeuAsnArgLeuPheArgAsn-----LysCys	686
DB	2530	GTCAAGGGTTGGCTATGTTTATTTCTGAATCGGCTACTTTTGGGAATCTGATATAGCAGCT	2589
QY	687	GluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluLutyrAlaGluThrLileLys	706
DB	2590	GAGTATAGACACCCCACTGTTCTCTCAGCTGTATAGAGAGAGGTATGAGAAAGCTTTGTTC	2649
QY	707	LysHisSerLeuGlnLysLileLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLys	726
DB	2650	AAGTTTACATTTGAAAGTTATTTGTTGTTGCTGTTTCTTGATATGCTAAAAATTCC	2709
QY	727	ArgIleValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLys	746
DB	2710	AGACTATTGATCATGATGCTCTGTCTCTCTGTAAGAATGCCGAATTCGAAGCTAGTAAA	2769
QY	747	AspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGlu	766
DB	2770	GAAATCTTTTGGCTTTTTCACAGAGATTTCTTAAGTGGTGAAGGTGACCTTTCCCGTCAC	2829
QY	767	LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr	786
DB	2830	CTTGGCTTATTGGGATTTACCTGTTTACCATGTTTCAGACACCATTTCATGAATTTGATTTT	2889
QY	787	AlaPheAsnAsnLeuLaValAspLeuArgAspGlyValArgLeuThrArgValValGlu	806
DB	2890	GCGGTTACAAATCTTGGCGTAGACTTGCATGTGGAGTGGCGCTTGTGGCAACCATGGAA	2949
QY	807	ValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeu	826
DB	2950	CTTCTCCACAGAACTGGGACCTCTCAAGAAACTCAGGATTCGGCAATAAGTCGTCTT	3009
QY	827	GlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyLutyrAlaAsnPheGlnLeu	846
DB	3010	CAAAAGATGACAAATGTTGCATTTCTTTCAGTTCTTAAATTCAGCAGGAATTAATTA	3069
QY	847	-----GlyGlyAspIleAlaLaGlnAspIleValAspGlyHisArgGluLys	862
DB	3070	AGTGATGAGCATGGAAATACAAATTCATCTAAGATATTGTGGATAGGCACAGAGAAAA	3129
QY	863	ThrLeuSerLeuLeuThrGlnLeuLileTyrLysPheArg-----	875
DB	3130	ACTCTCAGGTTGCTTTGGAAAAATAGCGTTTGTCTTTTTCAGGTGGATATTTCCCTTA	3189
QY	875	-----	875
DB	3190	GATCAATTAAAGGAAGAAATATGCTTTCTTAAACACACAAAGAGTATAAGAAAAACA	3249
QY	875	-----	875





```

CDS
/ds_xref="MIM:605481"
127..3555
/codon_start=1
/product="ASP protein"
/protein_id="AAH34607.1"
/db_xref="GI:21961301"
/db_xref="LocusID:259266"
/translation="MSTRAYTARCLNRLRAACRLFTSEKMKVAKIKKLEIEIEBARL
IVKDRHLKDVGRQKVLNWSYNFLMIRIGLETTYGELISLEDNSDVTGLAMFIL
NRLWNPDIAEYRHPYPHLYRDGHEALSFTLKKLLGLVLCFLDFAKAKSLRIDHP
CLFCKDAFFASKEIILAFSGEDLSRHGILGLPVMHVQTFDFBDFAVTNL
AVDLCQGVRIWMLLTONWDLKSLRIIPALISLQNMHNDVIYVLOKSGRIELSD
HGNTLSKDIWDRHRETKLRLMKIAFAFDVDSINLDQKEELAFKHTKSIKKTIS
LLSCHEDLNNKKKGDSPFOYSNKLMDVNAVCAFYNNKVKVSPFSDGR
VLCYLHHYHPFPAICORTQTVECTGTGSVINSSSESDDSLDMSLKAFDHE
NLSLYKELLENNKFNHVRSAVDRLDGGIPAMINSSSNTIIPDEKVIYIYLSFLCA
RLDLREIEARALIQTWRKYKLTDLKRHOEKAARIIOAVINFLAKORURKV
NAALVIQYWRALQKRLMKKEKLEKQNKAAASIOAMWRYRAKYLCKVKAAC
KIOWYRCWRAHKEYLATLKAVKIIOGCFYTKLERTFLNVRASAIIOQRWAILPA
KIAHEFLMKRHEAACLIQAHVRYGKRGVFLRKSAALIIOKVIRARBAKHRIK
YIEFKSTVILQALVRGWLVRKLEQRAKIRLHLFTAAAYHNLNVRIOAPAVLYLA
VKNANKQVNSICIQWRARLQKRFIOKTHSEKLEHGEQECLOSORNAASVIOXA
VRHFLRKROKFTSGIKIOALWRGYSWRKNDCTKIKATRLSLQVNMIREENKL
YKFTALALHYLLYKLSAILLEALKHLEVTVLSPCCENNAQSGAKSKIPVLIRCN
RSIPCMVIRYAVOVLLNVSKEYTKTGAVDVENCIDILLELQIYREKPKGVADKG
GSIPTKCLLAILLKTNNRASDVRSRKKVDRIYSYLYKLTAKHKMNTERILYKQKK
NSSISIPPIETPVTRIVSRSLKPDVLRDMMEEIINPLQAIQWMDTLGIPY"
1216..1395
/misc_feature
/note="CH; Region: Calponin homology domain"
/db_xref="CDD:smart00033"
2155..2769
/misc_feature
/note="COG5022; Region: Myosin heavy chain [Cytoskeleton]"
/db_xref="CDD:COG5022"

ORIGIN
Alignment Scores:
Pred. No.: 4,746-43 Length: 3760
Score: 932.50 Matches: 367
Percent Similarity: 40.63% Conservative: 229
Best Local Similarity: 25.02% Mismatches: 477
Query Match: 9.80% Indels: 395
DB: 9 Gaps: 51

US-09-914-698-1 (1-1861) x BC034607 (1-3760)
Qy 500 ILeAsnProAspPropheAlaalaThrThrThrIleAspPropheLeuAlaSerThrMet 519
Db 7 TTAACCCCTGACTTCTCACTGTAAAAACAAATATTTCTGAAGTAAATGCTGCTACTCTT 66
Qy 520 TyrLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysThrLeuAsnAla 539
Db 67 CTTTGTGGAAATAGAG-----AATCAACATATA----- 93
Qy 540 LeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLys 559
Db 94 ---ATAAGTGTCTCTAGA----- 108
Qy 560 LeuPheAsnGluValArgAsnLysGluLeuValAlaProThrLysGluGlnSer 579
Db 109 -----GCACCTTACAAAAGAGAAATGTCT 132
Qy 580 MetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu 598
Db 133 CTCAGAGCTTATCTGCTCGGTAGTGTAAACAGACTACGTCGTCGAGCATCGCGTGTG 192
Qy 599 PhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGln 618
Db 193 TTTACTTCTGAAAAAATGGTTAAAGCTATTAAAAAGCTTGAATTCGAATTTGAAGCTAGG 252
Qy 619 AlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIle 638
Db 253 CGGTTAATGTTTCGAAAAGATAGACACCTTATGGAAGATGTGGGAGAACGTCAGAAAGTC 312

```

```

639 LeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPhe 658
Db 313 CTGAATTGGCTGTGTCTCTCAATCCCTTTGGCTTCGAATGGTCTAGACCAACTTAT 372
Qy 659 GlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu 678
Db 373 GGAGAACTCATATCTTTGGAGATACAGTGAATGTACAGGGTTGGGTATGTTTATTCTG 432
Qy 679 AsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAlaTyr 695
Db 433 AATCCGCTACTTTTGGAACTCTGATATAGCAGCTGATATAGACACCCACCTGTTCCCTC 492
Qy 696 ThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPhe 715
Db 493 CTGTATAGAGATGCTCATGAAGAGCTTTGTCCAAAGTTTACATTGAAAAAGTTATTGTTG 552
Qy 716 LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeu 735
Db 553 TTGCTCTGTTTCTTGTGATTATGCTTAAATTTCCAGACTCATTTGATCATGATCCTTGTCT 612
Qy 736 PheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGlu 755
Db 613 TTCGTAAAGATGCCGAATTCAGCTAGTAAAGAAATCCTTTGGCTTTTTCAGGAT 672
Qy 756 LeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGln 775
Db 673 TTCCTAAGTGGTGAAGGTGACCTTTCCGTCACCTTGGCTTTATTGGGATTAACCTTTAAC 732
Qy 776 HisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeuAlaValAspLeu 795
Db 733 CATGTTCAGACACCATTTGATGAATTTGCGGTTCACAAATCTTTCGCGTAGACTTG 792
Qy 796 ArgAspGlyValArgLeuThrArgValValGluValLeuLeuLeuArgAspAspLeuThr 815
Db 793 CAATGTGGAGTGCCTTGTCCGAACCATGAACTCTCTCACACAGAACTGGGACCTCTCA 852
Qy 816 ArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAla 835
Db 853 AAGAAACTCAGGATTCGGCAATTAAGTCGTCTTCAAAAGATGCACAAATGTTGACATTGT 912
Qy 836 LeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIleAla 851
Db 913 CTTCAAGTCTTAAATCACGAGGAATTAAGTAAAGTGAAGCATGGAAAAATACAATCTTA 972
Qy 852 AlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTrpGlnLeuIle 871
Db 973 TCTAAGGATATTGGATAGCACAGAAAAAACTCTCAGGTTGCTTTGGAAAAATAGCG 1032
Qy 872 TyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGlnLysTrpTrpArg 891
Db 1033 TTTGCTTTTCAG----- 1044
Qy 892 ArgHisTrpLeuHisValValIleGlnArgIleArgHisLysGluLeuMetArgArg 911
Db 1044 ----- 1044
Qy 912 HisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrVal 931
Db 1045 -----GTGATATTTCCCTTAACTTAGATCAATTAAGAGAGAAAT 1086
Qy 932 LysLeuPheLys-ThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArg 951
Db 1087 GCCTTTCTTAAACACACAAAAGA----- 1108
Qy 951 gTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTr 971
Db 1109 ---GTATAAGAAAAACAA---TATCTCTACTAUCATGCCATTTTGTGATC----- 1153
Qy 971 pTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAl 991
Db 1154 -----TTATTATAAGAAAAAGCAAGGATAGTGGTTCTTCTTCAACAATAT 1203
Qy 991 aAla-----IlePheLeuGlnArgIleTrpArg 1000

```





```

1665 lLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerG1 1685
1685 lLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerG1 1685
2937 CTTAAACACCTAGAGTAGTTACTAGA----- 2964
1685 uPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLy 1705
2965 -----TTGCTCCACTTGTGTGAGACATGGCCAGAGTGAGCAATTCCTAAATATT 3020
1705 sGlnLeuIleGluArgCysSerArg----- 1713
3021 TGTTCGATCGAAGTTGTAATCGAGTAATCTCTTGATGGAAGTCATCAGATATGCTGT 3080
1714 ----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGluG1 1732
3081 GCAAGTCTTGTGTAATGTATTAAGTATGAGAAACTACTTCACAGAGTTTAAGTAGAGA 3140
1732 yGlyLeuValThrIleAlaGlnMetLeuLeu----- 1743
3141 AAATTGTATAGATATACTATTGGAGCTTTTCAGATATACCGAGAAAGCCTGGTAATAA 3200
1743 gTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAl 1763
3201 AGTTGCAGACAAAGCGGAGCAATTTTACAAAACCTTGTGTTGTTGGCTATTATTACT 3260
1763 aHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGlnAlaIleTy 1783
3261 G-----AAGACAAACAAATAGAGCTCTGATGACGAGTAGAGTCCAAAGTTGTGA 3311
1783 rMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaAr 1803
3312 CCGTATTACAGTCTCTCAAACTTACAGCTCATATAACATAAATG-----AATACTGA 3365
1803 gLysProProMetThrSerGlyArgTyrLysSerGlnLys----- 1818
3366 AAGAATA-----CTTTACAAGCAAAAGAGAAATCTCTTATAAGCAT 3407
1818 eAsnPheThrProCysSer-----LeuProSerLeuGluProAspPh 1832
3408 TCCCTTTTATCCAGAAACACCTGTGAAGCAGCAAGATAGTTTCAAGACTTAAGCCAGATTG 3467
1832 eGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAspThrI1 1852
3468 GGTTTTGAGAGAGATACATAGGAGAAATCAAAATCCCTGCAAGCTATTCAAATGGT 3527
1852 eLeuCysLysLeuGlnIle 1858
3528 GATGGATACGCTTGGCATT 3546

RESULT 9
AX876627 2453 bp DNA linear PAT 17-DEC-2003
LOCUS
DEFINITION
Sequence 11532 from Patent EP1074617.
ACCESSION
AX876627
VERSION
AX876627.1 GI:40031363
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 11532 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2453
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
876..>2453
/note="unnamed protein product"
CDS

```

```

/codon_start=1
/protein_id="CAE922264.1"
/db_xref="GI:40031364"
/translation="MELLTONWLSKRLIPAIRSLQKMNVDIVLQVLKSRGIELSD
EHNNTLSKDIVDRHREKTLRLWKIAFAQVDISLNDOLKEBIAFLKHTKSLTKTI
SLUSCHDDDLINKKGRDGSSEFQYSENKILAMDVNAVAFVKNKVFNTVFSFDG
RVLYLHHVPCYVPLDAILCQRTTQVTECTQTGVSVINSSSDSSLDMSLKAPDH
ENTSEYKEHLEPENKFNFLVRSADVRLDGGIPAMINSDMSTTIPDSKVVTIYLSFLC
ARLDLRKEITRAARLIQTWRKYKLKTLDRHQEREAARIIQLAVINFLAKRLKRR
VNAALVIOKVRRLVAOKLMLKKEKLEKVAASLIQGYWRYSYTRORFLKRY
SIILQSEIRMIIVATSYKRYLWATVTIQRHWRVLRKQDQORYEMKLSSTLIQSMF
RKWKORKMQSQVATVILQRAFWHLRKOAKENSALIIQSWYRMHKLKRYIIRS
CVVILIQKRFRCFQAKLY"

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 9,39e-38 Length: 2453
Score: 834.00 Matches: 235
Percent Similarity: 44.65% Conservative: 153
Best Local Similarity: 27.04% Mismatches: 228
Query Match: 8.77% Indels: 254
DB: 6 Gaps: 22

US-09-914-698-1 (1-1861) x AX876627 (1-2453)
QY 522 AspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuVal 541
DB 6 GATGAACGCTGGAAGGAAAGACAGGAGCAGGCGCTTCACTTGGTGGTTAAATTTTATATTA 65
QY 542 SerIleProAlaAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560
DB 66 ACC---CCTGATGATCTTCACCTGTGTAACAAATATTTCTGAAGTAATGCTGCTACTCTT 122
QY 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577
DB 123 CTTTGGGAATAGAGATCAACATAAATAAGTGTTCCTAGAGCACCTTACAAAAGAGGAA 182
QY 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaVal 596
DB 183 ATGTCTCTCAGAGCTTATCTGCTCGGTAGTGTAAACAGACTACGTGTCGACATGC 242
QY 597 GluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616
DB 243 CGTTTGTCTTACTCTGAAAAAATGTTTAAAGCTATTAAAAAGCTTGAATTTGAAATTGAA 302
QY 617 LysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArg 636
DB 303 GCTAGCGGTAAATTGTTGCGAAAGATAGACACCTATGGAAGATGTGGGAGAACGTCAG 362
QY 637 ThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluVal 656
DB 363 AAAAGTCCTGAATTGGCTGTTGCTTACAACTCTTTGTGCTTCGAATTTGGTCTAGAGACA 422
QY 657 ValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676
DB 423 ACTTATGGAGAACTCATATCTTTGGAAGATAACAGTAAGTGCACAGGTTGCTATGTTT 482
QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693
DB 483 ATTCGATCGCTACTTCTTGGATCTCTGATATAGCAGCTGAGTATAGACACCCCATGTT 542
QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle 713
DB 543 CCTCACCTGTATAGAGATGTCATCAAGAAAGCTTTGTCCAGAGTTTACATTTGAAAGTTA 602
QY 714 LeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733
DB 603 TTGTTGTTGGTCTGTTTCTTGATGATGCTAAATTTCCAGACTCATTCATGATGATCCT 662
QY 734 CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuArgPheSer 753
DB 663 TGTCTCTTCTGTAAGAGATGCCGAATTCAGAGGCTAGTAAAGAAATCCTTTAGGCTTTTCA 722

```





PN	JP	2002191363-A/11072
PD	09-JUL-2002	
PD	28-JUL-2000	JP 2000280990
PI	TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU	
PI	SAITO,	
PI	JUNICHI YAWAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,	
PI	KEIICHI NAGAI, TETSUJI OTSUKI	
PC		
	10, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC	
	12, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC	
	Primer for synthesizing full-length cDNA and use thereof FH	Key
FT	CDS	Location/Qualifiers
		(876)..(2453).
FEATURES		
source		Location/Qualifiers
		1..2453
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
ORIGIN		
Alignment Scores:		
Pred. No.:	9.39e-38	Length: 2453
Score:	834.00	Matches: 235
Percent Similarity:	44.65%	Conservative: 153
Best Local Similarity:	27.04%	Mismatches: 228
Query Match:	8.77%	Indels: 254
DB:	6	Gaps: 22
US-09-914-698-1 (1-1861) x BD156229 (1-2453)		
Qy	522	AspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaVal 541
Db	6	GATCAACGCTGGAAGGAAACAGCAGGACACAGGGCTTCACTTGGTGGTTAAATTTTATATTA 65
Qy	542	SerIleProLysAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560
Db	66	ACC---CCTGATGACTTCACTGTGTAAAAACAATATTTCTGAAGTAAATGCTGCTACTCTT 122
Qy	561	PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577
Db	123	CTTTTGGGAATAGAGAAATCAACAAATAAGTGTTCCTAGAGCACCTTACAAAAGAGAA 182
Qy	578	GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaVal 596
Db	183	ATGTCCTCAGAGCTTATCTACTGCTGGTGTAGGTAAACACACTACGTCGTGCAGATGC 242
Qy	597	GluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616
Db	243	CGTTTGTGTTACTTCTGAAAAAATGGTTAAAGCTATTAATAAAGCTTGAAATTTGAAATTGAA 302
Qy	617	LysGlnAlaIleArgIleArgSerAspArgAsnIleHisLeuAspValValMetClnArg 636
Db	303	GCTAGCGGGTTAATGTGTGAAAAAGATAGACACCTATCGAAAGATGTGGGAGAACGTGAG 362
Qy	637	ThrIleLeuGluLeuLeuLysCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluVal 656
Db	363	AAAGCTCGAATTCGGCTGTTCCTACAAATCCTTTGTGGCTTCGAATTTGGTCTAGAGACA 422
Qy	657	ValPheGlyGluLysIleGlnMetClnSerAsnArgAspIleValGlyLeuSerThrPhe 676
Db	423	ACTTATGGAGAACTCATATCTTTTGGAGATAAACAAGTGTGATGATGTACACGGGTGGCTATGTT 482
Qy	677	IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693
Db	483	ATTCTGAATCCCTACTTTTGGAAATCCTGAAATAGCAGCTGAGTATAGACACCCCACTGTT 542
Qy	694	AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle 713
Db	543	CCTCACCTGTATAGAGTGGTCATGAAGAAGCTTTGTCCAGTTTACATTTGAAAGATTA 602
Qy	714	LeuPheLeuLeuProPheLeuAspGlnAlalysGlnLysArgIleValLysHisAsnPro 733

```

Qy 915 aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLys 935
Db 1742 TCGACTCATCAACAAACAGTGGAGAAATATAAATAACATAAAACAGATCTCAAAACGCCATCA 1801
Qy 935 sThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaG1 955
Db 1802 G---GAGAGAGAGAAAGCTGCAGAAATATTCTCAATGGCTGTAATCAATTTCTTAGCAAA 1858
Qy 955 nLysGlnLeuTyrGlnSerTyrHisSerIleLeuThrIleGlnArgTyrTrpArgAlaG1 975
Db 1859 ACAAGATTGAGAAAGAGTAAATGAGCAGCACTGCTCATTCAGAAATATTGG----- 1910
Qy 975 nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLe 995
Db 1910 ----- 1910
Qy 995 uGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015
Db 1911 -----CGAAGAGCTTAGCAGAGAAATATTAAATGTTAAAGGA 1954
Qy 1015 aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035
Db 1955 AAAGCTGAGAAAGTTCAAAATAA-----GCAGCATCACTTATCAGGATATTGGAG 2008
Qy 1035 gThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055
Db 2009 AAGATTATTC-----ACTAGACAAAGATTTCTGAAATTTGAAATATTATTCAAT 2056
Qy 1055 tPheValGlnArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1 1075
Db 2057 CATCTGCAATCTAGATAAGA-----ATGATAATTGCTGTACATCTATATA 2104
Qy 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetAr 1095
Db 2105 ACGATATCTTTGGCTACAGTTACAAATCAGAGGCAATGGCGTCTATTATAAGAGAAA 2164
Qy 1095 GlyCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAla---Ty 1114
Db 2165 ACAAGATCAACAAAGATATGAAATCTTAAATCTTAAATCTTAAATCTTAAATCTATGTT 2224
Qy 1114 tArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134
Db 2225 CAGAAAA-----TGGAGCAACGTAAA-- 2246
Qy 1134 nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgG1 1153
Db 2247 -----ATGCAATCACAAGTAAAGAGCTACAGTAATATTGCAAGAGAGCTTTTAGAGA 2296
Qy 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169
Db 2297 ATGGCATTTAAGAAAACAGATTAAGAGAAAATTTCTGCTATTATATACATCATGTTA 2356
Qy 1169 rArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysI1 1189
Db 2357 TAGATGTCATAAGAAATTA-----CGGAAGTATATTATATTAGATCTTGCT 2404
Qy 1189 eIleAsnValGlnArgArgTrpArg 1197
Db 2405 TGTATTCATTCAGAAAGAGATTTCCG 2429

RESULT 11
LOCUS AK001380
DEFINITION Homo sapiens cDNA FLJ10518 fis, clone NT2RP2000814.
ACCESSION AK001380
VERSION AK001380.1 GI:7022604
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Negai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masubo, Y., Ninomiya, K. and Iwayanagi, T.
NEO human cDNA sequencing project
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 2453)
AUTHORS Isogai, T. and Otsuki, T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
source 1..2453
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2000814"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP2"
/notes="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
ORIGIN
Alignment Scores:
Pred. No.: 9,398-38 Length: 2453
Score: 834.00 Matches: 235
Percent Similarity: 44.65% Conservative: 153
Best Local Similarity: 27.04% Mismatches: 228
Query Match: 8.77% Indels: 254
DB: 22 Gaps: 22
US-09-914-698-1 (1-1861) x AK001380 (1-2453)
Qy 522 AspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysTyrLeuAsnAlaVal 541
Db 6 GATGAACGCTGGAAGGAAAGAGCAGGACAGGCGTTCACCTGGTGGTTAAATTTATATTA 65
Qy 542 SerIleProAlaAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560
Db 66 ACC---CCTGATGACTTCACCTGTAATAAACAATAATTTCTGAAGTAAATGCTGCTACTCT 122
Qy 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577
Db 123 CTTTGGGAATAGAGAAATCAACATATAAATAGTGTTCCTAGAGCACCTACAAAGAGGA 182
Qy 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaVal 596
Db 183 ATGTCCTCAGAGCTTATACCTGCTGGTGTAGTTTAAACAGACTACGTCGTCAGCATGC 242
Qy 597 GluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616
Db 243 CGTTTGTGTTTACTCTGAGAAATATGTTTAAAGCTTATTAAGAGCTTGAATTTGAAATTTGAA 302
Qy 617 LysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArg 636
Db 303 GCTAGCGGTAAATTTGTTGAAAGATAGACACCTATGGAAGATGTTGGAGAGAACGTAG 362
Qy 637 ThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuVal 656
Db 363 AAAGTCCTGAATTTGGCTGTTCCTACATCTTTTGTGGCTTCGAATTTGGTCTAGAGACA 422

```

QY 657 ValPheGlyGluIysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676  
 Db 423 ACTATGAGAGACTCATATCTTTGGAGAGATACAGTGATGTCACAGGTTGGCTATGTTT 482  
 QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693  
 Db 483 ATTCTGAATCGCCTACTTTTGGAAATCCTGATATAGCAGCTGATATAGACACCCACACTGTT 542  
 QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysIysHisSerLeuGlnLysIle 713  
 Db 543 CCTCACCTGTATAGATGTCATGAAGAAGACTTTGTCCAAAGTTTACATTCAGAAAGTTA 602  
 QY 714 LeuPheLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733  
 Db 603 TTGTTGTGTCGTCTTCTTGATATGCTAAATTTCCAGACATCATGATCATGATCCT 662  
 QY 734 CysLeuPheValLysIysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 753  
 Db 663 TGTCTCTTCTGTAAGATGCCGAATTCAGGCTAGTAAAGAAATCCTTTAGGCTTTTCA 722  
 QY 754 SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrVal 773  
 Db 723 CGAGATTTCTTAAGTGTGAAGTGCACCTTTCCCGTCGCCTTTGCTTATTGGGATACCT 782  
 QY 774 LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal 793  
 Db 783 GTTAACCATGTTTACAGACCATTTGATGATTTGATTTTGGCGTTACAAATCTTGGCGTAA 842  
 QY 794 AspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAsp 813  
 Db 843 GACTTGCATGTGAGTGCCTTGTGCGAACCACTTCTCCACACAGAACTGCGGAC 902  
 QY 814 LeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLys 833  
 Db 903 CTCTCAAGAAACTCAGATTCGGCAATAGTCTCTTCAAGAGATGCACATGTTGAC 962  
 QY 834 LeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAsp 849  
 Db 963 ATTGTTCTTCAAGTCTTAAATCACGAGGAATTTGAATTAAGTGATGATGAGAAATACA 1022  
 QY 850 IleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuThrGln 869  
 Db 1023 ATTCTATCTAAGGATATGTTGGATAGGCACAGAGAAAAAACTCTCAGGTTGCTTTGGAAA 1082  
 QY 870 LeuIleTyrLysPheArg----- 875  
 Db 1083 ATAGCGTTTGTCTTTCAGTGGATATTTCCCTTAACCTTAGATCAATTAAGAGAGAAATTT 1142  
 QY 875 ----- 875  
 Db 1143 GCCTTTCTAAACACACAAAGAGTATAAGAAACAAATATCTCTACTATCATGCCATTCT 1202  
 QY 876 -----SerProLysPheHis----- 880  
 Db 1203 GATGATCTTATTATAAGAAAAAGGCAAAAGGATAGTGTCTCTTTTGAACAATATAGT 1262  
 QY 881 -----AlaAlaAla-ThrValLeuGlnLys 888  
 Db 1263 GAAAAACAATAAGTTATGATGGATGGGTAAATGCTGTTTGTGCTTCTTATAATAAAG 1322  
 QY 888 sTrp----- 889  
 Db 1323 GTGAGAAATTTACAGTGTCTTTTCTCAGACGGCGGTGTTATGTTTACCTGATCCACCAT 1382  
 QY 889 ----- 889  
 Db 1383 TACCATCTTGTCTATGTGCCATTAGACGCTATATGTTCAGCGTACTACTCAAACTGTGGAA 1442  
 QY 890 -----Tr 890  
 Db 1443 TGTACGCAAACTGGTTCAGTGGTATTAAATTCATCATCTGAACTGATGACAGTTCTCTG 1502

QY 890 pArgArgHisTrpLeuHisValVal----- 898  
 Db 1503 GATATGTCATCTTAAAGCATTTTGATCATGAAATACTTTCAGAGCTATACAAAGAGCTCCTA 1562  
 QY 899 -----IleGlnArgArgIleArgHis----- 905  
 Db 1563 GAAAAATGAAAAAGAAAAATTTT-CACTTGGTTAGGTCGTCAGTTAGAGACCTTGGTGAAT 1621  
 QY 905 ----- 905  
 Db 1622 ACCTGCTATGATTAATCAATTCAGATATGTCAATATACAAATCCAGATGAAAAGGTGGTTAT 1681  
 QY 906 -----LysGluLeuMetArgArgHisArgAlaAl----- 915  
 Db 1682 TACCTATTTGTCATCTTCTTGTGCAAGGCTTTTGGATCTTCGTAAAGAAATAGAGCTGC 1741  
 QY 915 aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLys 935  
 Db 1742 TCGACTCATACAAACACACATGAGAGAAATATAAACTAAAAACACAGATCTCAAAOCGCCATCA 1801  
 QY 935 sThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaG1 955  
 Db 1802 G--GAGAGAGAGAGCTGCAAGATTTATCAATTCGCTGTGTAATCAATTTCTAGCAAA 1858  
 QY 955 nLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaG1 975  
 Db 1859 ACAAGATTGAGAAAAAGAGCTTAATGCAGCACTCGTCAATTCAGAAATATTGG----- 1910  
 QY 975 nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaIlePheLe 995  
 Db 1910 ----- 1910  
 QY 995 uGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaAlaGluThrAl 1015  
 Db 1911 -----CGAAGAGCTTAGCACAGAGAAATATTAAATGTATAAAGGA 1954  
 QY 1015 aArgLeuGlnArgSerGlnLysGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035  
 Db 1955 AAAGCTGGAATAAGTTCAAAATAAA-----GCAGCATCACTTATTTCAGGGATATTCGAG 2008  
 QY 1035 gThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055  
 Db 2009 AAGATATATCC-----ACTAGACAAAGATTTCTGAAATTTGAAATATTATTCAAT 2056  
 QY 1055 tPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1 1075  
 Db 2057 CATCTGCAATCTAGGATAAGA-----ATGATAATTGCTGTTTACATCTTATAA 2104  
 QY 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetAr 1095  
 Db 2105 ACGATATCTTTGGCTACAGTTACAAATTCAGAGGCATTCGCGTCTTATTTAAGAAGAA 2164  
 QY 1095 gLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAla---Ty 1114  
 Db 2165 ACAAGATCAACAAAGATATGAAATGCTTAAATCATCAACTCTTATATCAATCTATGTT 2224  
 QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
 Db 2225 CAGAAA-----TGGAGCAACAGTAAA-- 2246  
 QY 1134 nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgG1 1153  
 Db 2247 -----ATGCAATCAACAGTAAAAAGCTACAGTAATATTGCAAAAGAGCTTTTGAAGA 2296  
 QY 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169  
 Db 2297 ATGCGCATTTAAGAAAAACAAGCTAAAGAGAAATTTCTGCTATTATCATACAAATCATCGTA 2356  
 QY 1169 rArgMetArgGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysI1 1189  
 Db 2357 TAGAATGCATAAAGAAATTA-----CGGAAGTATATTTATATTAGATCTTCTGT 2404  
 QY 1189 eIleAsnValGlnArgArgTrpArg 1197





Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gregoire,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kwis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorenshewa,L., Louisleged,H., Lozano,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mapua,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwackelmech,O., Okwundu,G., Olarinpoosagon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,D., Primus,E., Pu,L.-L.,  
 Plopper,F., Poindexter,A., Popovic,D., Reeves,K., Regier,M.A., Reigh,R.,  
 Puazo,M., Quiroz,J., Rachin,E., Reuter,M., Richards,S., Riggs,F.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Rose,R., Ruiz,S.J.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstein,G. and Gibbs,R.A.

# TITLE JOURNAL REFERENCE AUTHORS JOURNAL

2 (bases 1 to 235179)  
 Worley,K.C.

Direct Submission  
 Submitted (09-OCT-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 235179)

Rat Genome Sequencing Consortium.

# TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24956330.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGMW

Center clone name: CH230-83E5

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 228193 bases at least Q40  
 Consensus quality: 230105 bases at least Q30  
 Consensus quality: 231450 bases at least Q20  
 Estimated insert size: 238014; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 4 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

\* 1 37618: contig of 37618 bp in length  
 \* 37619 37718: gap of unknown length  
 \* 37719 228862: contig of 191144 bp in length  
 \* 228863 228962: gap of unknown length  
 \* 228963 232645: contig of 3683 bp in length  
 \* 232646 232745: gap of unknown length  
 \* 232746 235179: contig of 2434 bp in length.

## FEATURES

source  
 1..235179  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-83E5"  
 1..1825  
 /note="wgs\_contig"

## misc\_feature

## ORIGIN

## Alignment Scores:

Pred. No.: 1.52e-30 Length: 235179  
 Score: 753.00 Matches: 274  
 Percent Similarity: 43.14% Conservative: 198  
 Best Local Similarity: 25.05% Mismatches: 324  
 Query Match: 7.91% Indels: 298  
 DB: 2 Gaps: 39

US-09-914-698-1 (1-1861) x AC0907078 (1-235179)

Qy 873 LysPheArgSerProlysisPheHisAlaAlaThrValLeuGln-----LysTrp 889  
 Db 159339 AGATTGAGA-----AAGCTGAATCTGATCCCTCTGTAATCCATCTGTTCAGACGG 159286  
 Qy 890 TrpArgArgHisTrpLeuHis-----ValValleGlnArgGile 903  
 Db 159285 TGGAGCGACGCAAGCTGCAGCTCCAAACCAAGCGCGGTGCATTCGACGACGCTTTT 159226  
 Qy 904 ArgHisLysGluLeuMetArg-----ArgHisArgAlaAlaThrValleGlnAlaVal 921  
 Db 159225 AGAATATGCGTTCGAGGAAACAGACAGAGAAAGCGCGCTGTGTGTATACAGTCGTGG 159166  
 Qy 922 PheArgGlyHis---GlnMetArgLysTrpValLysLeuPheLysThrGluArgThrGln 940  
 Db 159165 TACAGAATCACAGGGAATTGAGAAGTACATTACATT-----AGGTCT 159121  
 Qy 941 AlaAlaIleLeuGlnLysPheThrArgArgTrpLeuAlaGlnLysGlnLeuTrpGln 960  
 Db 159120 TGTGTTGTAGTATTCAGAGAAAGAGTTCGTGCTCTTCAGCGCCCAAGATGATGTAAAGAGA 159061  
 Qy 961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
 Db 159060 AGGAAAGACGCTGCTCTGACTCTCAGAGAGCACTTCAGGGCCGCTCAGATGGGAAAGTTA 159001  
 Qy 981 HisArgGlnArgPheValGlnLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArg 1000  
 Db 159000 GCGCGACTGACTATTTTCGAGAGCGCGCTGCATTCACACTCAG----- 158953  
 Qy 1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020



Db 158952 -----GCTGCCTTCAGGGGCGATGAAGGCTCGCCGTTCA 158920  
Qy 1021 GlnLysGlnGln---AlaAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeu 1039  
Db 158919 TACGTTTGCATAATCAAGCTGCCTGTGTCTTCAGTCTCTATTTGGAGAATGAGACAG--- 158863  
Qy 1040 GlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuMetPheValGlnArg 1059  
Db 158862 GGAAGAGTT-----AAATTCCTTAAATCTTAAAGAGAGCTGTATCAAAATTCGAAGCA 158812  
Qy 1060 ArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArg 1079  
Db 158811 CACATAAGAAAAACAC-----CAACAACACTACAGAAATATCAGAGATAAGAAA 158764  
Qy 1080 AlaAlaAlaAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAla 1099  
Db 158763 GCAGCTGTTCACAAATTCAGACTCATTTCCGAGCTTCATTTCCAGCCAGGAAGAGTCTAGCA 158704  
Qy 1100 AspTyrLeuAlaLeuArgSerSerValLeuLysValGln---AlaTyrArg----- 1115  
Db 158703 TCTTACCAGAAACACAGCTCTTCAGTCATTTGTCTACAGTCTGCATATACAGGGATGCNA 158644  
Qy 1116 -----LysAlaThrIleGln 1120  
Db 158643 GCCAGGAAAAACATTCGTCGTCTTTAACATCTGTAGTAAAGATTTCAGTTCATGTTATCGG 158584  
Qy 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
Db 158583 GCTTATATCTCCAGAAAGAGTTTCGAGCTTGAGGAATGCTGCAGTA----- 158536  
Qy 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
Db 158535 AAGCTGCAGTCAATTCGTAATGAAACAGTACAGTAAGCAGTATTTACAGTAAAGGCC 158476  
Qy 1161 AlaSerIleLeuValGlnLysArgTyrArg-----MetArgGlnGlnMetIleGlnAsp 1178  
Db 158475 GCTGCGGTATTTACAGCGGGTGGTACCCTGCCAAAAGAGAAAGAGTCTGACAGGTG 158416  
Qy 1179 ArgAsnAlaTyrLeuArgThr-----ArgLysCysIleIleAsnValGlnArg 1194  
Db 158415 CGGAAAGCTGTATCAAACTGCAGTCGTACTTCAGAGGATGCTGGTCAGAAAGCAGATG 158356  
Qy 1195 ArgTrpArgAlaThr-----LeuGlnMetArgGluArg 1206  
Db 158355 AGCTCCAAAGCAAGCTGCCATTTCTCTGCAGTCTTACTTCAGGATGTGGAGAGTGAGG 158296  
Qy 1207 LysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLys 1226  
Db 158295 CAGCGTACTCTGAAGCGGCGCAAGCGGCTCTTGTCTATTCACACTTCTACAGTGCCTAC 158236  
Qy 1227 ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVal 1246  
Db 158235 CGAGCACAAAGTCCATCAGAGGAAGATCTTCTCTCAGGTCAAAGGAGCAGCTGTTGTTTG 158176  
Qy 1247 Gln-----LysArgArgArgAlaLeuLeuGlnMetArgLys----- 1258  
Db 158175 CAAGCAGCTTACAGAGCGCGCAGAGCCCGGGAATCAAGCAGCAGTCTACAGCTGCC 158116  
Qy 1259 -----GluArgGlnGluTyrLeuHisLeuArg 1267  
Db 158115 GTTACAATTCAGAGGTTTTTAGAGGCCACAGTCTCAGAGGATGAATACCAGACTGTGCTT 158056  
Qy 1268 GluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArg 1287  
Db 158055 CAGCTGTGTGAAGTGTGAGAGTGTACAGGGGACACAGAGTTGCTTCTGACATGAGA 157996  
Qy 1288 AlaLysTyrArgGlyThrGlnAlaAlaVal-----SerCys----- 1299  
Db 157995 ATTCACCTTTCTAAAGACAAGGGAGGCTGTGTGTGTTCTGCTCTGCTGCTGCTGCTGG 157936  
Qy 1299 ----- 1299  
Db 157935 CAAGTTCGGCAGCAGCTCAGGAGGCAACATCAGGCTGCAGTGAAGATCCAGGCTCTGTTTC 157876

Qy 1300 -----LeuGln 1301  
Db 157875 AGAATGGCCAGGGCCAGAGGCGAGTATAAACTAGTGAGAAATGCTGCAGTCTCATCCAG 157816  
Qy 1302 MethIstPrArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArg 1321  
Db 157815 CAGCATGTTCAGGGCAGCGGCTGCAGAGAGAGGCGACCTCCGAGTACATCCCGCTCCGT 157756  
Qy 1322 GlnAlaIleThrLeuGlnArgArgTyrArgAlaArgLeu----- 1335  
Db 157755 CACGCCGCCCTGGTGTTCAGGCTTACCTGGAAGGAAGATGCTGAGGAGAGATCCGA 157696  
Qy 1336 -----AsnMetIleLysGlnLeuLys 1342  
Db 157695 CAGCGCACCAGTCCGCGCGCCCTCATACAGTCGTACTACAGAATGTACAGTCCAGCTAGG 157636  
Qy 1343 SerTyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLys 1362  
Db 157635 AAGTGGAGCTTCATGAATAACCGCTGCTCTGCAGATTTCAGTGTGTATAGGGCTACAGA 157576  
Qy 1363 AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleLysValGln 1382  
Db 157575 GTTGGAAAGGAACAGCGCCATTTGTACTTGAACAACAAAGGCGCTGTAGTAAATTTACAG 157516  
Qy 1383 ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIle----- 1395  
Db 157515 TCAGCTTACCCTGGT-----ATGAAGAGCGAGAAAGACAGATAACTGAATGCCACAAAGCAGCA 157459  
Qy 1396 -----GluValTyrGlnLysGln-----Arg 1402  
Db 157458 GTCACTATACAGTCTAAATTTAGGGCTACAGAACACAGAAAGAAATATACACCTGTAGA 157399  
Qy 1403 GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuLysLys 1422  
Db 157398 ACTTCTCTCTATTAAATTCAGAGATGGTATCGGAATATTAAATATACCACCCAGCAACGT 157339  
Qy 1423 AlaGlyTyrArgArgIleArgLeuSerLeuSerIleGlnArgLysTrpArgAlaThr 1442  
Db 157338 GAGCATACCTTACCTTAAAGAGAGCAGCAACTAAAGTCCAACTGTGTACAGAGGC--- 157282  
Qy 1443 ValGlnAlaArgArg----- 1447  
Db 157281 ATTCAGTAAGGAGACACATTCAGCACATGTACATGCGCGCCACACTCATTTAAAGCCTTG 157222  
Qy 1447 ----- 1447  
Db 157221 TTTAAGATACAGCAGTCGAGCAGCGGCTGTGATCCAGGTGAGGTACAGAGCGTATTAT 157162  
Qy 1448 -----GlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
Db 157161 CTAGGTAAATTCAGCAGGAAAGATGACTGACGACTTTTAGGGCTATTAAAGACCTTCAA 157102  
Qy 1464 AlaPheIleArg-----AlaThr 1469  
Db 157101 GCCGTGTACAGAGCAGCAGAGTTAGAACACTTTAAGAAAGATGATTCGCGAGCAAG 157042  
Qy 1470 LeuLeuMet-----ArgGlnGlnArgGluPheGluMetLysArgArg 1484  
Db 157041 CTCATTTCAGTCATCTTTAGGAGACACAGACAGCAACATATTTCCACCGCTTGAGGAAG 156982  
Qy 1485 AlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGln 1504  
Db 156981 GCAGCCACAGTGTGACGAGAGTCCCGAGCAGTAAAGAGGAAGCTCTCAGTTTCAG 156922  
Qy 1505 AspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPheArg----- 1521  
Db 156921 AGGTACAGCAGACTGAGCGCTTCTGTAACTCTGCTCCAGCTGCTTTAGGGGAGTAA 156862  
Qy 1522 AlaAsnArgSerMetLysGln----- 1528  
Db 156861 GCGCGGAGACACTTAAAGACTATGCATTTAGTCTGCTACTGTCTCATTTCAGAGGAGATTACGG 156802



```

QY 1529 -----AlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGln 1545
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156801 ACACACTGATGAGAGGAGTTCCTCTCTCAGGAAACAGTCAATTGGATTGAGAGA 156742
QY 1546 LysPheArgGlyLysLeuMetIleGluGlnArgAsnCysPheGln-----LeuLeu 1563
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156741 CAGTACCGTGCAAGA-----CTTCAGCGCAAGTATTGCGCAACAATTAAGTGTG 156691
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156690 GAGAGCGCGTTTACAAAGATCCAGTCATCATCAAAAGGATGATGTAAGAAAGTGTG 156631
QY 1584 GlnAlaLeu-----MetThrProGluMet 1591
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156630 CAGAGATGCACAGGCGCTGCTACTGTGATTTCAGGCCACTTTCGGAATGCACAGAGCTGT 156571
QY 1592 MetAspLeuLeuArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyr 1611
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156570 GTGAGGTACAGCGTTTGAAGAGGCGCTCAGTTGTGATCCAGAAAGCAATACCGAGCACAC 156511
QY 1612 LeuIleArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGln 1631
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156510 -----AGAACTGCAGAACTGCAGAGCGACGCTTTTGTTCAGGCAGACAAAGCTGCT 156460
QY 1632 LeuArgGlnGluAla-----LysAlaValAsnSerValArgCysLysVal 1646
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156459 CTGACCTTCAGGCTGCCCTTTAGGGCGTGAAGCCAGAAACCAATTTGAAGACTATGCAT 156400
QY 1647 GlnGluAla-----ValArgPheLeuArgGlyArgPheIle 1658
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156399 GCTTCTGCACACCTTATTTCAGAGTAATTTAGAGCTAGATAAGTAAAGGACAAATTCCTT 156340
QY 1659 AlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHis 1678
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156339 GCTTCTCAGAAAGCTGCTATATTGCGTCAG-----AGAAATACAGACCAACCTTTAT 156286
QY 1679 LeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAla 1698
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156285 GCCAAGCATAAATGACCAATCTTGCGAGTTACGAAAGCAGCCATTACATACAGTCC 156226
QY 1699 Ile-----ArgSerGluValAspLysGlnLeuIleGluArg 1710
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156225 TCTTTTCGAAGACTGATGCTCAAAAGACGATTACAAAGAAAGG 156184

RESULT 14
AY365047
LOCUS Pan troglodytes ASPM protein (ASPM) gene, complete cds.
ACCESSION AY365047
VERSION AY365047.1 GI:37953276
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE Pan troglodytes
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 62965)
AUTHORS Kouprina,N
TITLE Pan troglodytes genomic DNA sequence of ASPM gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 62965)
AUTHORS Kouprina,N
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2003) LBC, NCI, 9000 Rockville Pike, Bethesda, MD
20892, USA
FEATURES
source
    1..62965
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        <137..>62239
        /gene="ASPM"
        join(<137..433,2447..2590,2737..4216,6675..6779,
11304..11450,12956..13201,14195..14262,16489..16630,
21493..21584..21669,
22199..22420,23935..24142,24229..24371,24487..24615,
28544..28738,41355..46109,50394..50560,52381..52477,
53302..53511,54505..54654,55523..55714,56176..56368,
56483..56637,58133..58309,59592..59761,62137..>62239)
        /gene="ASPM"
        /product="ASPM protein"
        join(137..433,2447..2590,2737..4216,6675..6779,
11304..11450,12956..13201,14195..14262,16489..16630,
21493..21584..21669,
22199..22420,23935..24142,24229..24371,24487..24615,
28544..28738,41355..46109,50394..50560,52381..52477,
53302..53511,54505..54654,55523..55714,56176..56368,
56483..56637,58133..58309,59592..59761,62137..62239)
        /gene="ASPM"
        /codon_start=1
        /product="ASPM protein"
        /protein_id="AA04850.1"
        /db_xref="GI:37953277"
        /translation="MANRVGRCWEVSTERRPAGLRGPAEAEASPPVLISHP
CRSPFLCDVLLGASRTLSALDNPNEVAEVKISHFPADLIGSVSRQFVLQPK
KIVISVNTPLKEGVREIMTFVNDVLKHQAILIGNAEBOCKKRSKWDITTKKKS
ASTSNRVNTQNTKTFVSQKVDVRVPLQACENLANNRGPTTENNLSILENK
IPISIPAFNEYHGATCLPLSVRSITVSSLSASRENELNVHVASNVKVSFNKAV
TETSPSNVNGSGENSKLSITPNVSTLTNTQSTHFLSPDSFVNNHSGANNELEL
VTCLSMPKDNKSPVHLESTIAHEIQKILSPDSFTKDNGLQDLSESVNPLS
PFQKLDNWAYMCTSCQCTCKPSNENQVPSQPEDWRKSEVSPRPECQGSKPAI
FEEELVMKSNYSYFKONNPFSAVODISSHNNKOPKRRPILSATVTKKATREN
QTEINPKAKRCLNSAEGHEKVINNOKEDEFHSYLPIDPILSKSKYKNEVTPSI
TTASVARRKSDSGMEDGNVRVAITEHTEVREIKRIHFSPSPKTSVAKTKNTPPI
SKRIINREKILKKKTDLSIPRTPLSKTNKTKPIIAVAQSLTPIKPLKTPTRHPM
PFAKMPYDERMEKEQSGFTWMLNFILTPDDFTVTNISEVNAATLLGTEHQHKI
SVPAKTEMSLRATACRLNRACRLTSEKMWAKKKUEIEIEARLRVRK
DRLHLDVGEORQKVLNLLSNPLWIRLETGTYGELISLESDNSVTKGLAMFLNRL
WNPDIAARYRHPVPLHYRDGHEVLSKFTLLKLLVCLFYDAIKSLRIDHDPCLFC
KDAEFKASKELILAFSRDPLSGEGLSHRLGLGLPVNHRVQTFPEFVAFNVAVDL
QCGEFLRTMELLTQNDLSKLRIPATISRLQKMNVDIVOLVKSRLGIELSDHGT
ILSKDIDVHRREKTLRLWKTAFAPOVDISLNDQLEKIEAFLKHTSKTKTISLSC
HSDLLNKVGGDCGSEFEQYSENKILMDWNAVCAFNNKVENFTVSFSDVILCY
LIHHYPCVPDAICQRTQCTVECTQGVSVNSESSESDSLMSDQPHDENTSE
LYKEILENEKKNFLHVRGAVRDGGLGIPAMINHSMSNTIPDEKVIITVLSFCALLD
LRKEIARAALIQTTWRKYLKTDLRKHOREKAARIIOAVINFLAKRLRKRVAAL
VIOKYRVLVAORLMLKLEKVNKAASLIOGYWRYSTROFLKLYYSILQ
SEIPMLIATVSYRYLVATVTOREHRAVLRKQDOQRYEMLSKSTLIQSMRQWQ
RMSQSVATVILQAFREHRLKQAKQNSAIIQSNWRMHEKRLKIIIRSCVLI
QRRFCQAKLYKKKKSILTIQKYAYLKGKERTNYLQKRAALIQLOAFARLK
AHNLQIRACVCIQSYWRMRQDRVFLNKKTIILQAHVRKHQQRKYKKKAAV
IIQTHFRAYIFAMKVLASQKIRSAVILQSAYRGAAALFIOQYKSKLTAOKRREYV
YVSKKEFLSNATIKQSIQVWKQTKQYLHRAAALFIOQYKSKLTAOKRREYV
QMBESCIQLOAFRGVLYRKOMLRKAVISLSQYPRMKAQOYLKMYKALIVIQNY
THAYAQVNQRNFIQVKKATCLOARVGYKVRQLIKQSSAALKIQSAFRGYNRV
KYQVLSLIQIQRVRAKTLHDTTRTHPLTKAAVILQSAYRGVKKVQRIRRHQA
ALKIQSAFRMAKQKQFRLFTAAVILQONFRWAGTKQRMETYLEHVALMLQSMW
KGLTURLQROHQKCAIIQSYRYMHVQCKWKIMKKAALLIQKYRAYISIGREONHL
YLKTKAAVTLQSAVGMKVRKRIKQCNKAAVTIOSKYRAYTKKAYATVRSALIIQ
BWYRGKLTNHQHEVLANKTAIKILOSRYGIRVRRLHQMHRATFIKAMKWHQS
RISYHTRKAAIVIQRCRAYIQGQREKILTIKAVQLQASFRGVRRLRQWQ
IATLTIQSNRYRYQTFNKLKTIQVQRYRAMEKNTIQQRYKRLHSHVIYIQA
IFRGKARHLKMMHIAATLQRRFTLMRRRFLSKLKTALIQRKYRLHSHVIYIQA
QFLOQNAVIKIQSSYRWRMIRKMRMHRATFIQATPRMHLMRVQALQKQASVVI
QOQYQANRAALQROHLYRHSVAVILQAPRGMTKRLHKSMSHSAATIQSFRSLI
VRRFISLKKATIYQRYKATICAHKHLOPHLRKAAITIOSYRYLWVKLQOEM
QRAVLIQATFRMHRTYITFQWKGASILIQOHTYTYRAKLRNRYIRQWHSVIO
AAYGMKARQLLRKHKASIIQSTYRYWYQFYKQLQWATKIIOEKYRANKKQKA
QRYRKLTAVRTQAVICIQSYRGFKVKDQNMHRAATLIQSFYRMRKAVDYETKK
FOHNEKKTQVQGFQDMNKQIQEQHQAIIIOQKCKAFKIRHILHRTVYSI
QAVVIOYRYLYRVKTKERNFLAVQKSVRTIOAAPRGMYRQKLVNVEKMAIIV
NKSALCCTRSKTOHEAVQSGEWMIOEWYKASGLACQSEAYHSQSRAAVTQKACRM
VTRKLETKQKALRIQFLONAVYRRFVQKRAATLIQHYPTWTRKQFLYLRKAA
VYLQNHRYAFISAKHQRYVLIQIRSSVIIQARSKGFIQKRFQIRKSTIKIQAWR
RYRKKYLCKVKAACKIQAWYRCWPAHEKYLKAVKIIQCFPTKLTFTFLNVR
SAIITQKWRALISAKIAHEHFLMKHRAACLIQAHVRYGKRGVFLRQKSAALIIO
KYIRAREAGKRIRKIVIEFKKSTVILQALVRGLVYRKRILEQRAKIRLILHFTAAAYIH

```

LNARVIAQAYKLYLAVNKANKOVNSVICIQWFRARLOEKRFIOKVHSIKKIBEGOB  
CLSORNEAAVIOKAYHFVLRKOKKFTSIIKIQALWEGYSRWKNDCTKIKARL  
SLOWNREIRENKRYKPTALALHYLLTYKHLNAILALEKHLVFRPLSLPCENWAO  
SGALSKIFLRNSCNBIRYPCWENIRYAVQVLLNVSKTEKTSIAVENCIDILDELL  
QIYREKPNVADKGSIFTKTCLLAILKTTNRASDVRSKVDRIYSLYKLTAH  
KRWNTIRILYKOKONSSISIPETPFTVRTRIVSRILKPDWILRDNMBEITNPLQAI  
QWMDTLGIEY"

ORIGIN

Alignment Scores:

Pred. No.: 7,87e-31 Length: 62965  
Score: 744.50 Matches: 286  
Percent Similarity: 41.29% Conservative: 207  
Best Local Similarity: 23.95% Mismatches: 379  
Query Match: 7.83% Indels: 322  
DB: 9 Gaps: 43

US-09-914-698-1 (1-1861) x AY365047 (1-62965)

Qy 868 TrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaIleThrValLeuGln 887  
Db 41655 TGGCATTTA-----AGAAACACAGCTAAAGAACAAATCTGCTATTATCAATCAA 41705  
Qy 888 LysTrpTrpArgHis-----TrpLeuHis-----Val 897  
Db 41706 TCATGTTATGAATGCATAAAGATTACGGAAATATATTATATTAGATCTTGTGTGTT 41765  
Qy 898 ValIleGlnArgGileArg-----HisLysGluLeuMetArgArgHisArg 913  
Db 41766 ATCATTCAGAAAGATTTCGGTGTCTTCAAGCCCAAAAGTTGTATAAAGAAAGAAAGAG 41825  
Qy 914 AlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeu 933  
Db 41826 TCATATCACTAACATCCAGAGAGTACTACAAGCATATCTGAAGAAAGATTGACGGCACC 41885  
Qy 934 PheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyrLeu 953  
Db 41886 AACTATTTCAGAAACAGAGCTGCAGCCATTCAATTACAAGCTGCTTTTAGGAGA---CTG 41942  
Qy 954 AlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArg 973  
Db 41943 AAAGCTCATTAATTATATAGACAAATTAGAGCTGCTTGTGTATTATTCAGTCTACTGGA 42002  
Qy 974 AlaGlnGln----- 976  
Db 42003 ATGAGACACAGACAGCTTCGATTTTAAACCTTAAGAGACTATTATCAAAATTCACGCA 42062  
Qy 977 -----LeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaIlePhe 994  
Db 42063 CATGTAAGAAACATCAACACGACAGCAAAATATAAGAAAGATGAAGAAAGCAGCTGTTATA 42122  
Qy 995 LeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThr 1014  
Db 42123 ATTCAGACTCATTTCCGAGCTTATATTTTGGCCATGAAGATTTAGCATCTTCACGAAA 42182  
Qy 1015 AlaArg-----LeuGlnArgSerGlnLys-----GlnGlnAlaAlaAlaSer 1028  
Db 42183 ATAGGCTCTGCTGTCATTGTGCTGCTGCATATAGAGGGAGTCAAGCCAGGAAATG 42242  
Qy 1029 TyrIleGln-----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGln 1043  
Db 42243 TATATTACATCTCATCTGTTTATAAAGATTCAATCAVATTATTCGTGCTTATGTTTCT 42302  
Qy 1044 ArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSer 1063  
Db 42303 AAAAGGAATTTTGGACCTAAARAAATGTCACAATA-----AAATTCAGTCA 42350  
Qy 1064 LysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsn 1083  
Db 42351 ATTGTAAAGTGAACAAACAGTAAACAATATTATTCATTTAAGCAGCTGCCTATT 42410  
Qy 1084 ILeGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAla 1103

Db 42411 ATCCAGCAATGTTTACCGTTCCAAAAAATAGTCGACAAAAAGAGAGAGTATATGCAG 42470  
Qy 1104 LeuArgSerValLeuLysValGlnAlaTyr----- 1114  
Db 42471 ATGCGGAATCTTGTATCAAACTGCAAGCATTTGTTAGAGATACCTTGTCCGAAAGCAG 42530  
Qy 1115 -----ArgLysAlaThrIle-----GlnMetArgIleAsp 1124  
Db 42531 ATGAGGTTACGAAGAAAGCTGTTATTTCACTACAGTCTTATTTCAGAAAGAGGCT 42590  
Qy 1125 ArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgAla 1144  
Db 42591 CAGCAGTATTATCTGAAAATGTATAAAGCAATTATTGCTCATCAGAAATTAATCATGCA 42650  
Qy 1145 IleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeu 1164  
Db 42651 TACAAAGCACAGGTCATATCAGAGGAAGAACTTCTTGAAGTCAAAAAGCAGCTACTTGC 42710  
Qy 1165 ValGlnLysArgTyrArg-----MetArgGlnGlnMetIle--- 1176  
Db 42711 TTGCAAGCAGCTTACAGAGGTTATAAAGTACGCCAGCTAATCAAAACAACAATCTATAGCT 42770  
Qy 1176 ----- 1176  
Db 42771 GCTCTTAAATTCAGTCTGCTCTTTTAGAGGCTATAATAAAGGTAAAAATATCAATCTGTG 42830  
Qy 1177 -----GlnAsp 1178  
Db 42831 CTTCAATCTATATAAAGATTACAGATGTGTACAGGCGGTACAGACTCTTCATGATACA 42890  
Qy 1179 ArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGln-----ArgArg 1195  
Db 42891 AGAACATATTTTGAAGACAAAGGCGCTGTGATTTCCCTCCAGTCTGCTTATCGTGGC 42950  
Qy 1196 TrpArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThr 1215  
Db 42951 TGGAAAGTTCCGAAACACAGATTAGAAGGGAACATCAAGCTCCCTTG----- 42995  
Qy 1216 ThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu 1235  
Db 42996 -----AAGATTTCAGTCTGCTTTTAGA-----ATGGCCAGGCCGCAAGAACAG 43037  
Qy 1236 PheLeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGln 1255  
Db 43038 TTTAGATTGTTTAAACACAGCAGCATTAGTCAATCCGCAAAATTTTCAGAGCATGAGACTCA 43097  
Qy 1256 MetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGln--- 1274  
Db 43098 GGAAGGAAGCAAGCTATGAGTATATTGAATCCGTCATGCGGTACTGATGCTTCAATCT 43157  
Qy 1275 -----AtgArgPheHisAlaGlnLysSer----- 1282  
Db 43158 ATGTGGAAGGAAAAACACTGAGAGAGCAGCTTCAAGGCAACATAATATGTGCTATCATC 43217  
Qy 1283 -----MetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAla 1296  
Db 43218 ATACAGTCATATAGATGTCATGCAACAAAGAAAGTGAATAATCATGAAAAAGCT 43277  
Qy 1297 ValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSer 1316  
Db 43278 GCTCTTCTGATTCAAAAGTATTATAGGGCTTATAGAGTATGGAAGAGACAGAAATCATTTA 43337  
Qy 1317 PheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsn 1336  
Db 43338 TATTGAAAACAAAGCAGCTGTGGTAACTTTTACAGTCAGCTATCTGTGT----- 43388  
Qy 1337 MetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThr 1356  
Db 43389 ---ATGAAAGTGAAGAAAGAAATAAGGATTGCAACAAAGCAGCAGCTCATCTATACAGTCT 43445  
Qy 1357 ArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGlu 1376  
Db 43446 AATACAGAGCTTACAAAAACCAAAAGAAA-----TATGCAACCTATAGAGCT 43493



```
/tissue type="human esophagus tumor"  
/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host  
DH10B_sites SfiIA + SfiIB"  
/dev_stage="adult"  
polyA_signal 5718. .5723  
polyA_site 5738  
ORIGIN
```

```
Alignment Scores:  
Pred. No.: 8.49e-32 Length: 5760  
Score: 737.00 Matches: 408  
Percent Similarity: 38.35% Conservative: 331  
Best Local Similarity: 21.17% Mismatches: 637  
Query Match: 7.75% Indels: 554  
DB: 9 Gaps: 77  
  
US-09-914-698-1 (1-1861) x HSM808955 (1-5760)  
Qy 253 HisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThrPro 272  
Db 451 CAATCAGAATTACTATCATGTCATA---CAAAGCACAGTCAATCAGAGGA-AGAACTTCT 506  
Qy 273 ThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArg 292  
Db 507 TGCAAGTCA-----AAAAGCAGCTACT-TGCTTGCAAGCAGCTTACAGA 550  
Qy 293 LysTyrSer-----GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyr 310  
Db 551 GGTTATAAAGTACCCAGCTAATCAACAACAATCTATAGCTGCTCTTAAATTCAGTCT 610  
Qy 311 AlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsn 330  
Db 611 GCTTTAGAGGCTATAAT-----AAAAGGTAATAATATCAATCTGTGCTTCAATCT 661  
Qy 331 -----ArgTyrTyrGlnGlnGlnGlnGlnIleLysAlaLysAspLeu 345  
Db 662 ATAATAAAGATTCAGAGATGGTACAGGCGTAC-----AAGACTCTT 703  
Qy 346 AsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeuPheAsnHis 365  
Db 704 CATGATCA-----AGAACACATTTTGTGAACACAAAG 736  
Qy 366 SerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGlyArgLysSerVal 385  
Db 737 GCAGCTGTGATTCCTCCAGTCTGCTATCGTCTGCTGGAAGGTT---CGGAACACAGATT 793  
Qy 386 LysGlySerProValLysAsnProHisLysArgArgSerHisGluLeuSer----- 402  
Db 794 -----AGAAAGGGAACATCAAGCTGCCTTGAAGATT 823  
Qy 403 -----PheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThr 417  
Db 824 CAGTCTGCTTTAGATGGCCAGGCCAGAACACAGTTTAGATTGTTTAAACACAGCAGCA 883  
Qy 418 ValAlaIle-----SerProLysLysLysGlnArgValGlu--- 429  
Db 884 TTAGTCATCCAGCAAAATTTACAGACATGGAATTCGAGGAAGCAATGATGAGATAT 943  
Qy 430 -----AspThrThrLeuPro 434  
Db 944 ATTGAATCCGCTCATGCGGTACTGCTGCTTCAATCTATGTGGAAGGGGAAACACTGAGA 1003  
Qy 435 ArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTyrProHis 454  
Db 1004 AGACAGCTTCAAAGGCAACATAAATGTCTATCATCATACAGTCATCATCTATAGATGCAT 1063  
Qy 455 AlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThr 474  
Db 1064 GTGCAACAAAAGAAAGTGGAAA-----ATCATCAAAAAGCTGCT--- 1102  
Qy 475 ProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyr----- 490  
Db 1103 -----CTTCTGATTCAAAGATATATATAGGCTTACAGATTGGGAAGA 1144
```

```
Qy 491 ---AspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThr 509  
Db 1145 GAACAGATCATTTATATTGAAACA-----AAAGCAGCTGTAGTA 1186  
Qy 510 ThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529  
Db 1187 ACTTTA----- 1192  
Qy 530 GlnAlaAspPheLysLysTyrLeuAsnAlaLeuValSerIleProAlaAspLeuAla 549  
Db 1193 CAGCAGCTTATCGT----- 1207  
Qy 550 AspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeu 569  
Db 1208 -----GGTATGAAAGTGAGAAAAGAAATAAAGATTGC---AACAAAGCAGCA 1252  
Qy 570 ValValAlaProThrLysGluGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGlu 589  
Db 1253 GTCACATATACAGTCTAATAATACAGAGCT-----TACAAAACCAAAAAGAAATATGCA 1303  
Qy 590 ThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSer 609  
Db 1304 ACCTATAGAGCTTCAGCTATATATAATT-----CAGAGATGGTATCGAGGTATT 1351  
Qy 610 LysValAlaValTyrValAsnLysGlnAlaLeuArgIle-ArgSerAspArg----- 626  
Db 1352 AAAATTACAAACCATCAGCATAGGAGTATCTTAAATTGAAGACAGACGCAATTAATAATC 1411  
Qy 627 -AsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeuCysPheAs 646  
Db 1412 CAATCTCTTTATAGAGTA-----TTAGAGTTAGAAGACATATTCAA 1453  
Qy 646 n-----ProLeuTyrLeuArgLeuGluValValPheGlyGluLy 661  
Db 1454 CACATGCACAGGCGAGCCACTTTTATTAAAGCCATGTTTAAATGCATCATCAGCAATA 1513  
Qy 661 sIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLe 681  
Db 1514 AGTTACCATACAATGAGAAAAGCAGCTATTG----- 1546  
Qy 681 uPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTy 701  
Db 1547 ATTCAAGTA-AGATGT-----AGAGCATATATTCAAGGTAAATGACGCGTGAAGATA 1599  
Qy 701 rAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAs 721  
Db 1600 CCTCACAATTTGAAAGCTGTAAAGTCCTTCAGGCAAGTTTT----- 1642  
Qy 721 pGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerPr 741  
Db 1643 -----AGAGGAGTAAGA-----GTTAGACGCACTCT 1668  
Qy 741 oHisLys---GluThrLysAspIleLeuLeu-----ArgPheSerSerG1 755  
Db 1669 TAGAAGATGACAGCTGCAGCAACACACTCATTCTAGTCAAACTACAGAGATACAGACAGCA 1728  
Qy 755 uLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuG1 775  
Db 1729 AACATACTTTAAT-----AAGTTAAAGAAATAACAAAACAGATACA 1770  
Qy 775 nHisArgGlnThrPheLeuAspGluPheAspTyrAlaPhe-----AsnAsnLeuAl 792  
Db 1771 GCAAGATATCTGGGCAATGAAAGAAAGAAACATACATAATTTCAAAGGTATACAACTGAG 1830  
Qy 792 aValAspLeu-----ArgAspGlyValAr 800  
Db 1831 GCATTCGTATATACATTCATTCAGGCTATTTTATAGGGGAAAGAAAGCTAGAGACATTAAA 1890  
Qy 800 gLeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVa 820  
Db 1891 AATGATGCATATACCCGCAACTCTCAATTCAGAGGAGATTTAGAACTCTAATGATGAGAAG 1950
```



Db 3985 CTGCAAGATTCAAGCCTGGTATAGATGTTGGAGAGCACACAAAGAAATATCTAGCTATATT 4044  
QY 1426 ----- 1426  
Db 4045 AAAAGCTGTTAAATTAATCAAGGTGCTTCTATACCAAACTAGAGAGAACACCGTTTTT 4104  
QY 1427 -ArgileLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaAr 1446  
Db 4105 GAATGTGAGAGCATCAGCAATATCATTCAGAGAAATGAGAGCTATACTTCTCGCAAA 4164  
QY 1446 gArgGlnArgGlnIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheII 1466  
Db 4165 GATAGCTCATGAACACACTCTTAATGATAAAAGA----- 4198  
QY 1466 eArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgAlaAl 1486  
Db 4199 -----CATCGAGCTGC 4209  
QY 1486 aValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTy 1506  
Db 4210 TTGTTTGATCCAAAGCACAATTATAGAGA-----TATAAGGAAGCGAGGTCTT 4257  
QY 1506 rGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526  
Db 4258 TCTTCGCGAGAACTCTGCTGCTTTGATCATACAAAATATATACGACCGAGGAGGCTGG 4317  
QY 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLy 1546  
Db 4318 AAAGCATGAAGGATAAAATATATTTGAATTTAAANAATCTACAGTTATCTCTACAGCACT 4377  
QY 1546 sPheArgGly-----LysArgLeuMetIleGlnGlnArgAsnCysPheGlnLeuLe 1563  
Db 4378 GGTGGGTGGTGGCTAGTACGAAAAAGATTTTATAGACAGAGAGCCCAAAATTCGACTTCT 4437  
QY 1563 uArg----- 1564  
Db 4438 TCACCTTCACTGCAGCTGCATATTATCACCTGAATGCTGTGTAGAAATTCAAAAGAGCCTATAA 4497  
QY 1565 -----CysSerMetPr 1568  
Db 4498 ACTTTACCTGGCTGTGAAGAAATGCTAACAGACGAGTTAATTCAGTCAATCTGTATTCAGAG 4557  
QY 1568 oGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe---GlnAlaLeuMe 1587  
Db 4558 ATGGTTTCGAGCAAGA-----TTACAAGAAAAGAGATTTATTTCAGAAATATCA 4605  
QY 1587 tThrProGluMetMetAsp-----LeuIleArgGlnLysArgAlaAl 1601  
Db 4606 TAGCATCAAAAAGATTGAGCATGAAGGTCAAGAAATGCTCTGAGCCAGCGAAATAGGGCTGC 4665  
QY 1601 aLysValIleGlnArgTyTrpArgGlyTyTrpLeuIleArgArgGlnLysHis----- 1619  
Db 4666 ATCAGTAATACAAAGACAGCATGGCCATTTTCTCCCTCGTAAAGACGAGGAAAAATTCAC 4725  
QY 1620 -GinglyLeuLeuAspIleArg----- 1626  
Db 4726 TAGTGAATCATTTAAATTCAGGCATTATGAGAGGCTATTCCTGGAGGAGAAAAATGA 4785  
QY 1627 ----LysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLy 1645  
Db 4786 TTGTAACAAAATTAAGCTATACGACTAAGTCTTCAAGTTGTTAATAGG-----GA 4836  
QY 1645 sValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaVa 1665  
Db 4837 GATTCGAGAGAAAACAAACTTACAAAAGAACTGCACCTTGCACTTCATTACCTTTTGAC 4896  
QY 1665 lLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGI 1685  
Db 4897 ATATAAGCACCTTCTGCCATTCTTGAGGCTTAAACACCTAGAGGAGTAGTTACTAGA-- 4954  
QY 1685 uPheMetSerThrPheCysTyTrpGlyIleMetAlaGlnAlaIleArgSerGluValAspLy 1705  
Db 4955 ----TTGTCTCCACTTTTGTGTGAACACATGGCCAGAGTGGAGCAATTTCTAAAAATATT 5010

QY 1705 sGlnLeuIleGluArgCysSerArg----- 1713  
Db 5011 TGTTTGTATCCGAAGTTGTTAATCGCGTATTCCTTGTATGAAGACTCATCGATATGCTGT 5070  
QY 1714 ----IleIleLeuAsnLeuAlaArgTyTrpAsnSerThrThrValAsnThrPheGlnGluGI 1732  
Db 5071 GCAAGTCTTCTTAATGTATCTAAGTATGAGAAAACCTACTTTCAGCAGCTTTATGATGTAGA 5130  
QY 1732 yGlyLeuValThrIleAlaGlnMetLeuLeu-----Ar 1743  
Db 5131 AAATGTATATAGATATACTATTGGAGCTTTTGCAGATATATCCGAGAAAAGCCTGGTAATAA 5190  
QY 1743 gTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAl 1763  
Db 5191 AGTTTCAGACAAAGCGGGAAGCAATTTTACAAAACCTTGTTGTTGTGGCTATTTTACT 5250  
QY 1763 aHisCysProLysLysArgLysIleIleHisAspTyTrpMetThrAsnProGluAlaIleTy 1783  
Db 5251 G-----AAGACAACAATAAGAGCCTCTGTATGTACGAAGTAGGTCCAAAGTTGTGA 5301  
QY 1783 rMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaAr 1803  
Db 5302 CCCTATTTCAGCTCCCTACAAACTTACAGCTCATAAACATAAAATG-----AATACTGA 5355  
QY 1803 gLysProProProMetThrSerGlyArgTyTrpLysSerGlnLys-----II 1818  
Db 5356 AAGAATA-----CTTTACAAGCAAAAAGAGAATTCCTTCTATAAGCAT 5397  
QY 1818 eAsnPheThrProCysSer-----LeuProSerLeuGluProAspPh 1832  
Db 5398 TCCTTTTATCCAGAAAACACCTCTGAAGACCAGCAATAGTTTCAAGACTTAAAGCCAGATTG 5457  
QY 1832 eGlyIleIleArgTySerProTyThrPheIleSerSerValTyTrpAlaPheAspThrII 1852  
Db 5458 GGTTTTGAAGAGATAACATGGAAGAAATACAAAATCCCTCTCAAGCTATTCAAATGGT 5517  
QY 1852 eLeuCysLysLeuGlnIle 1858  
Db 5518 GATGGATACGCTTGGCAAT 5536

Search completed: September 15, 2004, 06:57:05  
Job time : 17563 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 14, 2004, 21:44:50 ; Search time 1427 Seconds  
(without alignments)  
5540.222 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSOME2  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPFO\_spool\_h/US0914698/runat\_10092004\_150004\_28194/app\_query.fasta\_1.2055  
-DB=N Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0914698 @Cgn\_1\_1\_1091@runat\_10092004\_150004\_28194 -NCFU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	9514	100.0	5586	3 AAA37761	Aaa37761 Drosophil
2	9482	99.7	6003	4 ABL11881	Ab111881 Drosophil
3	9386.5	98.7	8330	4 ABL11880	Ab111880 Drosophil
4	834	8.8	2453	4 AAH14237	Aah14237 Human cDN
5	707.5	7.4	3887	4 AAH14236	Aah14236 Human cDN
6	700.5	7.4	4027	6 ABQ99318	Abq99318 Human cod
7	542.5	5.7	1115	5 ABV20287	Abv20287 Human pro
8	542.5	5.7	1115	5 ABV20247	Abv20247 Human pro

C 9	542.5	5.7	1115	5	ABV26078	Human pro
C 10	542.5	5.7	1115	5	ABV20220	Human pro
C 11	542.5	5.7	1115	5	ABV26049	Human pro
C 12	542.5	5.7	1115	5	ABV26118	Human pro
13	439.5	4.6	752	5	ABV09513	Human pro
14	437	4.6	2414	4	AAH14270	Human cDN
15	437	4.6	2414	7	ACC50994	Human bla
16	353	3.7	576	5	ABV30689	Human pro
17	353	3.7	576	5	ABV39657	Human pro
18	342	3.6	24789	4	ABL28640	Drosophil
C 19	332	3.5	802	5	ABV25946	Human pro
C 20	332	3.5	802	5	ABV20113	Human pro
21	322	3.4	7677	5	ABV23205	Human pro
22	322	3.4	7677	5	ABV29042	Human pro
23	322	3.4	14800	6	ABL66291	lung canc
24	317.5	3.3	15231	6	ABK63598	Abk63598 Rat seque
25	317.5	3.3	15231	7	ABT42078	Toxicity
26	317.5	3.3	15231	9	ADB59205	Toxicity
27	317.5	3.3	15231	9	ADB53026	Primary r
28	316.5	3.3	14835	6	AAS94858	Human DNA
29	316	3.3	771	4	AAH05820	Human cDN
30	316	3.3	6401	7	ACC50236	Breast ca
31	303.5	3.2	8673	4	ABL11109	Ab111109 Drosophil
32	300.5	3.2	476	5	ABV39583	Human pro
33	300.5	3.2	476	5	ABV30748	Human pro
34	300.5	3.2	476	5	ABV39716	Human pro
35	300.5	3.2	506	5	ABV09572	Human pro
36	294	3.1	9551	2	AAZ22301	cDNA enco
37	293	3.1	5229	8	ADB83244	Human cDN
C 38	286	3.0	11200	4	ABL11108	Drosophil
39	285.5	3.0	501	5	ABV09438	Human pro
40	285.5	3.0	5928	4	ABL10555	Drosophil
41	282.5	3.0	6946	5	ABV25339	Human pro
42	279.5	2.9	461	5	ABV00403	Human pro
C 43	279.5	2.9	8527	4	AAI60039	Human pol
44	278.5	2.9	5230	7	AAAD49464	Human ves
45	277	2.9	5068	5	AAF32508	Human mal

ALIGNMENTS

RESULT 1  
AAA37761  
ID AAA37761 standard; DNA; 5586 BP.  
XX  
AC AAA37761;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DB Drosophila Asp coding sequence.  
XX  
KW Asp; Drosophila; microtubule organising centre; MTOC; mitosis inhibitor;  
KW tumour cell; ds.  
XX  
OS Drosophila sp.  
XX  
FH Key  
FT CDS Location/Qualifiers  
FT .5586  
FT /\*tag= a  
FT /product= "Asp"  
XX  
PN WO200052478-A1.  
XX  
PD 08-SEP-2000.  
XX  
PF 03-MAR-2000; 2000WO-GB000785.  
PR 04-MAR-1999; 99GB-00005007.  
XX  
PA (UYDU-) UNIV DUDEER.  
XX  
PI Glover DM, Avides MDC;  
XX



DR WPI; 2000-594203/56.  
 DR P-PSDB; AAY90350.  
 XX Use of Drosophila Asp polypeptide for identifying substances capable of  
 PT disrupting microtubule organization center integrity and use of the  
 PT identified substances for inhibiting mitosis in tumor cell.  
 XX  
 PS Disclosure; Page 44-47; 51pp; English.  
 XX  
 CC This sequence encodes the Drosophila Asp protein. The invention relates  
 CC to the use of Drosophila Asp polypeptide (or its homologue, or fragment)  
 CC capable of stimulating formation and/or maintenance of microtubule  
 CC organising centres (MTOCs), in an assay for identifying a substance  
 CC capable of disrupting MTOC integrity. Asp polypeptide or its homolog is  
 CC useful for identifying a substance capable of disrupting MTOC integrity.  
 CC Substances identified by the method can be used to inhibit mitosis, e.g.  
 CC in tumour cells  
 XX  
 SQ Sequence 5586 BP; 1532 A; 1466 C; 1387 G; 1201 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 5586  
 Score: 9514.00 Matches: 1861  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0

US-09-914-698-1 (1-1861) x AAA37761 (1-5586)

QY	1	MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeu	20
DB	1	ATGAGCTAGTGTGGAGCCCGCTTGAGTGGCTCGAAGGAGACGCTGCAGCTAATA	60
QY	21	AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro	40
DB	61	GACAAACGGCAACTCCGAAAGAGGTGATGATCATCTCAAGTCCAAAGAGCAACCGCG	120
QY	41	ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr	60
DB	121	GTCAAGNACCCGGCAAAATTTCTACTGTGGCAAGACCTGTGAGCTGAAATGCGCGACA	180
QY	61	GlyAlaGlyLysThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMet	80
DB	181	GGAGCTGGCAAGCAATGAAAGCGTGTATCCGCTCTGTGAGCAAGCAAAAGAGCGCATG	240
QY	81	SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro	100
DB	241	TCGTGACGACAGCGCTCCCTCCAGCAGACATGGCGAGTGACTGTCTCTGCGGTCCC	300
QY	101	AlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLysThr	120
DB	301	CTGTGATGGGCATATCCACCTCCACAGGCTCTCTTGTGAGAGAGAAATGTATACAGACT	360
QY	121	ProGlnGluGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer	140
DB	361	CCACAAGAGAGCGGTATACATATACACAGCTCCAGCCTCGCAGCTCTTAAGGAAATCTAAGC	420
QY	141	ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThrProLeuThr	160
DB	421	CCATGACGCGAGAAACCTACTCGAGCTGATTGACAAATCTGGCAATTCACACCTCTCACC	480
QY	161	GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr	180
DB	481	GAAACCCGTGGCAAGACAGACTACCATTTTCCCGACAAATCTGGCAGCTGGCCACA	540
QY	181	ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThr	200
DB	541	CCAAACACTTAAGGGAATGAAATCATGTGCTAATGATATGCGCGCGCTGCAATCACT	600
QY	201	ProAspAsnLeuGluAsnGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu	220
DB	601	CCAGATGATCTAGAAGATCAGCCCTGCCCAAAACAAACGTTTCGATGTAAAGCATTCGAG	660
QY	221	ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr	240
DB	661	ACTATCAATATTCGTGGACACCTTGGACTGCTCCAGGATCGATGGACAAACCGCATACG	720
QY	241	ProLeuAsnLysThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle	260
DB	721	CCCTAAATAAGACAACAACCATTTGTGATGCCAGCACACACAGAGCTCTGGCTGTATT	780
QY	261	HisGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu	280
DB	781	CATGAGGAGGAGGACCAAGTCCACTTAGACCCCAAGAGAGCGCATACAGACCTG	840
QY	281	LysArgAspIleLysLeuValGlySerProLeuArgLysTyrSerGluSerMetLysAsp	300
DB	841	AAGAGGGACATTAAGTTGGTGGTTCCACCTTACGAAAGATTCCGAGTCCATGAAAGAT	900
QY	301	LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnLysSerMetProAsnLeuAsn	320
DB	901	TTGTCTCACTTCTATCGCCACAACCTAAGATATGCCATTCAAGGGTCTATGCCCTAATCTAAAT	960
QY	321	GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlnGlnIleGlnIle	340
DB	961	GAAATGAAATCCGCTCGATCGAACAGATAGATATTACCAGGAGCAGCAGATCCAGATA	1020
QY	341	LysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu	360
DB	1021	AAGGCCAAGACTTGAATAGTCTCTAGTAGGAGGCTAGTTTGGCCGCGCAGCAGAG	1080
QY	361	PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluVal	380
DB	1081	TTTCTATTCAACACACAGTGCATCTCGCTCAGTCCAGTCTGTTTAACTCTCCATGAAGTA	1140
QY	381	GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu	400
DB	1141	GGTCGGAAGTCGTGAGGGAAGTCGGTGAAGATCCTCACAGGCCCTCTCTCATGAG	1200
QY	401	LeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThrValAlaIle	420
DB	1201	TTGAGTTTTCGAGTGCACCTAGCAACGAATCATTTGACCGCAATGAATCTGATGCCATT	1260
QY	421	SerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla	440
DB	1261	TCCTCTCTTAAAGAACACCGGTTGAGGACACTACTCTGCCCAAGTGCAGCGCCGCA	1320
QY	441	AsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPhe	460
DB	1321	AATGCATCTCAAGAGACGATGATGCCACGCTGGCCACACGCCCAATCCAGAAAGTTT	1380
QY	461	LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp	480
DB	1381	AAGCTAGCACAAACCATGTCACTGATGAAGAGCCGCCACACCAAGAAAGTCAGGGAC	1440
QY	481	ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle	500
DB	1441	ACTAGCATTCAGCTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	1500
QY	501	AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr	520
DB	1501	AACCCGATCCATTTGACGACACTACGACAAATGATCCATTTCTGGCATCTACCATGTAT	1560
QY	521	LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysThrLeuAsnAlaLeu	540
DB	1561	TTGGATGAACAGCGCTGGGATCGTCATCAAGCTGACTTTTAAAGAGTGGTAAATGCCCTT	1620
QY	541	ValSerIleProAlaAspIleAspAlaAspLeuAsnLysIleAspValGlyLysLeu	560
DB	1621	GTCTCCATACCCGCTGACCTGGACGCGAGATTAAATAACAAATAGACGTTGGTAAGCTG	1680
QY	561	PheAsnGluValArgAsnLysGluValValAlaProThrLysGluGlnSerMet	580
DB	1681	TTTAAAGAGTGGCCAAACAAAGAGCTCGTGTGGTCTCCACCAAGGAGGAGCAGCTCATG	1740



QY 581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600  
 Db 1741 AACTACCTAAACGAATACCGCTGAGAGCGCTTGGTAAGCGGCTGTGGAGCTCTCTTC 1800  
 QY 601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
 Db 1801 AGTGAGCAGATCGCGCTGCATGCTCCAAAGTGGCGGTATATGTCAACAAGCAAGCTCTG 1860  
 QY 621 ArgGleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
 Db 1861 CGATCCGAGGATCGTAATCTTACCTAGACGATGTTATGCAACGACCATCTTGTAG 1920  
 QY 641 LeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlu 660  
 Db 1921 CTGCTGCTTGTCTCAATCCCTTTGGCTGGCTTGGACTGGAAGTGTCTTTGGCGAG 1980  
 QY 661 LysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
 Db 1981 AAGATCCAGATCGAGTCTAATCGAGACATTTGAGGCTTCAGCACCTTTATCTCAATCGC 2040  
 QY 681 LeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
 Db 2041 TTGTTCCGGAATAGTGTGAGGAGCAGAGGTACAGCAAGCGCATACACTCCACGAAGAG 2100  
 QY 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720  
 Db 2101 TACGGGAGACCATTAAGAGACATCATTTGCAGAAAATCTCTTTTGTCTTCTCTC 2160  
 QY 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
 Db 2161 GATCAAGCTTAAGCAGAAGCGCATCGTCAAGCAATCCCTGTTTGTGTTTAAAAAGTGG 2220  
 QY 741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerLeuLeuLeuAlaAsnIle 760  
 Db 2221 CCACATAAGAGACCAAGATATCTGCTGCCCTTCTGCTGGAGCTGCTCGCCCAACAT 2280  
 QY 761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
 Db 2281 GGTGATATTACGGGGAATCTGCTGCCCTGGCTACGTTCTACAGCACCGCCAAACATTT 2340  
 QY 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
 Db 2341 TTGAGCAGAGTTCCGATATAGCTTCAACAACTTGGCTGTGGACTTAAGAGATGGCGTGAGA 2400  
 QY 801 LeuThrArgValValGluValIleLeuLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
 Db 2401 CTACCCGAGTCTGGAGGATAATTTTACTACCGGATGATCTAACCCGCGAGTTAAGGGTG 2460  
 QY 821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyValAlaLeuGly 840  
 Db 2461 CCAGCCATCTCTCGCTTCAGCGGATCTTCAATGTAAAGCTAGCTCTGGCGCACCTTGGT 2520  
 QY 841 GluAlaAsnPheGlnLeuGlyAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
 Db 2521 GAAGCCAACTTCCAGCTAGCGCGCACATCGCGCCCAAGACATCGTTCAAGGACATCGT 2580  
 QY 861 GluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
 Db 2581 GAGAAAGCGCTTCCCTGCTCGGCAACTATTTACAAATTCGCTCGCCCAAGTTTCAT 2640  
 QY 881 AlaAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGln 900  
 Db 2641 GCGCGGCCACCGTCTCCAGAAATGGTGGCCGCTCACTGCTGCGAGCTGTGTATCCAG 2700  
 QY 901 ArgArgIleArgHisLysGlnLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
 Db 2701 CGTGCATTCGCGCAAAAGAGCTTATGGCTCGCCACCGGGCGCTACTGTCTATTGAGGCC 2760  
 QY 921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
 Db 2761 GGTCTCCGGGCCACACAGATGAGAAGTACGTGAAGTTGTTTAAACCGGAACGCACTCAG 2820  
 QY 941 AlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGln 960

Db 2821 GCCCGCAATAATCTCCAAAAGTTCCCGTCCGTTATTTGGCCCGAGAGCAGTATATCATG 2880  
 QY 961 SerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGln 980  
 Db 2881 AGCTATCATCAGTATTATCACCATCATCGCGCTGGTGGAGCCCAACAACTGGGAAGGCGAG 2940  
 QY 981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTyrArg 1000  
 Db 2941 CACCCCGCAGCGGTTTGTGAGCTCCGAGAGCTGCGATCTTTCTCAGCGGATCTGCGCG 3000  
 QY 1001 ArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
 Db 3001 CGACGACTCTTTGCCAAAACACTATTGGCGGGCGGGAACAGCCAGACTTCCAGCGATCG 3060  
 QY 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGly 1040  
 Db 3061 CAAAACCAACAGCGAGCTGCTAGTTATATTTCAATGCAATGGCGAACGTTATCAGCTGGGC 3120  
 QY 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
 Db 3121 AGAATTCAGCGACACGAGTTCCTGCGCAGAGGACCTCATCATCTTTGTTTCAGGCGCAG 3180  
 QY 1061 MetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
 Db 3181 ATGCGAGCAAGTGGAGTATGCTGGAGCAGCGCAGAGGAGTTCAGCAACTAAAGCGTGCA 3240  
 QY 1081 AlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
 Db 3241 GCTATAAATATCCAAACACGCTGGCGAGCGAAGCTTTCAATGAGAAAGTGCAACGCTGAT 3300  
 QY 1101 TyrLeuAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
 Db 3301 TATTTGGCACCTTCGTTCCAGCGCTTCTTAAAGTTCAGGCTTACAGAAAGCCCAATCCAG 3360  
 QY 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
 Db 3361 ATGAGATAGATCGTAATCAGTACTATTTCCTGCGAAAAAATGTTATCTGCTCTGCAACAG 3420  
 QY 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
 Db 3421 CGACTGAGGGCCATCATGAAAATGCGGAACAGAGGGGAAAATTAATCTCAGGCTGCGAAAT 3480  
 QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
 Db 3481 GCTTCAATACCTAGTTCAAAAACGCTACCGCATGCGTCAACAAATGATCCAGATGAAAT 3540  
 QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
 Db 3541 GCATATTTAAGAACCCGCAATGTATCATCAATGTTCCAGAGCGCTGGAGAGCCACTCTG 3600  
 QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnInThrThrLysArgIleGln 1220  
 Db 3601 CAGATGCTCGAGAAAGGAAGAACTTACCTTCATCTCCAAACGACAAACAAACGAATTCAA 3660  
 QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
 Db 3661 ATCAAGTTCCTGCGCAGCGGAAATGAAAGCAAGAGCGGAGTTTCTTTCAGCTGAAA 3720  
 QY 1241 LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 Db 3721 AAGTTACCTTGTAGTCCAGAAAGCTCGCGGGCTTGTCTGCAGATCGGAAAGAGAGCGC 3780  
 QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
 Db 3781 CAGAGTACCTTACATCTGCGGAGGTTACCATTAAGCTGCGAGCGAGATTTCATGCTCAA 3840  
 QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
 Db 3841 AAAACAATGCGGTTTCATGCGAGCCCAAGTACCGCGCACCCAGGCTGCTGTGTGAGCTGCTG 3900  
 QY 1301 GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320

Db 3901 CAGATGCATTGGCGTAATCATCTCTTTAGAAACGGGAGAGGAACAGTTTCTTACAACTG 3960  
Qy 1321 AtcGlnAlaAlaIleThrLeuGlnArgArqTyrArgAlaIleLeuMetIleLysGln 1340  
Db 3961 CGTCAAGCAGCAATAACACTGCGAGCGACCGGTACCGAGCTCGTCTGAATATGATCAAGCAG 4020  
Qy 1341 LeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAla 1360  
Db 4021 TTGAAGATTACGCCAGCTGAACAGGACAGCTATTACCAATTCAAACCCGATATAGAGCC 4080  
Qy 1361 LysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGlnAlaIleLys 1380  
Db 4081 AAGAAGCAATGCAAAAGCAGTGGTCTTGTACCAAAAGCAAGGGAAGCAATATCAAA 4140  
Qy 1381 ValGlnArgArgTyrArgLysGlnLeuMetArgLysGlnIleGluValTyrGlnLys 1400  
Db 4141 GTCAACGACGATACCGCGCAATCTGGAGATGAGGAGAGATGAAAGTCTACCAAAA 4200  
Qy 1401 GlnArgGlnAlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeu 1420  
Db 4201 CAGCGCCAGGACGTCATCCGCTTTCAGAAATGTGGCGCAGTATACGCGACATGCGGCTG 4260  
Qy 1421 CysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArg 1440  
Db 4261 TGCAAAAGCGGCTACCGAAGGATTTCGACTCAGTTTCATTGAGCATTTCAACGCAAGTGGCGG 4320  
Qy 1441 AlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArg 1460  
Db 4321 GCCACGTGCGGCTGTCGCCAACGGGAGATCTTTAAGCACCATCCGCAAGTGGCGA 4380  
Qy 1461 LeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGlu 1480  
Db 4381 CTTATGCGAGCAATTTATCAGAGCAACTTTTACTGATGCTCAGCAGCGCAGGGAATTCGAG 4440  
Qy 1481 MetLysArgArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeu 1500  
Db 4441 ATGAAGCGAAGGGCTGCGGTAGTGATCCAGCGCGGTTTCGTGCTCGCTGTCTATGCTA 4500  
Qy 1501 LysAlaArgGlnAspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPhe 1520  
Db 4501 AAGCGAGACAGATTACCAATTAATCCATCTCTGTGATCTCTGTGTCGCGCGCAATTC 4560  
Qy 1521 ArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAla 1540  
Db 4561 CGTGTAAATCCGACGATGAAGCAGGCGACCGCAGGAATTTCTCAGCTGCGTACTATTGCA 4620  
Qy 1541 ValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGlnGlnArgAsnCysPhe 1560  
Db 4621 GTCCATTTGCAACAAAAGTTCCGTGGCAAGCGTCTAATGATTGAGCAGCGTAATTTGTTT 4680  
Qy 1561 GlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArg 1580  
Db 4681 CAATGCTCGCTGTTCCATGCCGGGTTTCAGGCAGTGCCTCCGCGCTTTATGGCTCGC 4740  
Qy 1581 LysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAla 1600  
Db 4741 AAACGATTCCAGGCCCTGATGACACCCGAGATGATGACCTCATCCGCCAGAAAGCGCGCC 4800  
Qy 1601 AlaLysValIleGlnArgTyrTrpArgGlyTyrIleLeuArgArgGlnLysHisGln 1620  
Db 4801 GCCAAGGTTATACAGCGTTACTCGCGGGGTATCTAATCCGACGACGCTCAGAACCCAG 4860  
Qy 1621 GlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGlnAlaLysAlaValAsn 1640  
Db 4861 GGTCTCTTGATATCCGAAGCGTATTGCTCAGCTGGCGCAAGAGGCAAGGCGGTAAAC 4920  
Qy 1641 SerValArgCysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheIleAlaSer 1660  
Db 4921 TCTGTGCGTGCAAAGTCCAGAGCGGTGGTTCCTTCGCGGACGCTTTATCGCATCA 4980  
Qy 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu 1680  
Db 4981 GATCGGTTAGCAGTGCCTAAGTCAATTGATCGTCTTTTCGCGCACTGTGCCACACCTGCTC 5040

Qy 1681 MetTyrCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg 1700  
Db 5041 ATGTGGTTCGAGTTCATGTCCACGTTTGTATGGCATCATGGCTCAGGCCATTCTGA 5100  
Qy 1701 SerGluValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArg 1720  
Db 5101 TCAGAGGTGATGAAGCAGCTTATAGAGCGCTGAGCGGATCATCCTTAATTTGGCCCGC 5160  
Qy 1721 TyrAsnSerThrThrValLeuThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMet 1740  
Db 5161 TACAATAGCACCAACGCGTGAACAGCTTCCAGAGGCGGTTTGGTCACCATGGCCAGATG 5220  
Qy 1741 LeuLeuArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyr 1760  
Db 5221 TTATTGGCTGGTGGCGCAAGACAGTGAATATCAACATTTGTGACCCCTCATTTGG 5280  
Qy 1761 ValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGlu 1780  
Db 5281 GTATTGCGCCACTGTCTCTAAAAAGCGAAAGATCATTCACGACTATATGACCAACCCAGAG 5340  
Qy 1781 AlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGln 1800  
Db 5341 GCCATTACATGTGCGGAACTAAGAGCTCGTGCCCGCAAGGAAGATGAAGCAA 5400  
Qy 1801 AsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPhe 1820  
Db 5401 AATGCAGCAAGCGCGCCCAATGACAAAGTGAGCGGTATTAAGAGCCAAAGATAAACTTC 5460  
Qy 1821 ThrProCysSerLeuProSerLeuGluProAspPheGlyIleIleArgTyrSerProTyr 1840  
Db 5461 ACGCCGTGTTCCCTGCGCCAGCTTGGAGCGGACTTCGGAATCATCCGCTACAGTCCCTAC 5520  
Qy 1841 ThrPheIleSerSerValTyrAlaPheAspThrIleLeuCysLysLeuGlnIleAspMet 1860  
Db 5521 ACGTATTATCTGTCGCTTACGCTTCGATACGATTTTGTGCAAGCTGCGAGATCGACATG 5580  
Qy 1861 Phe 1861  
Db 5581 TTT 5583  
RESULT 2  
ABL11881  
ID ABL11881 standard; cDNA; 6003 BP.  
XX AC ABL11881;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30125.  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30125.  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30125.  
KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 30125.  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 30125.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PSDB; ABB67778.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 XX interactions.

PS Claim 1; SEQ ID NO 30125; 21pp + Sequence Listing; English.

XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL161176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 6003 BP; 1629 A; 1566 C; 1502 G; 1306 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.: 0 Length: 6003  
 Score: 9482.00 Matches: 1855  
 Percent Similarity: 99.89% Conservative: 4  
 Best Local Similarity: 99.68% Mismatches: 2  
 Query Match: 99.66% Indels: 0  
 DB: 4 Gaps: 0

US-09-914-698-1 (1-1861) x ABL111881 (1-6003)

QY 1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuLeu 20  
 DB 418 ATGAGCTAGTGTGGAGCCCGCTCTTCCAGGTGGCTGCAAGGAGCGCTGCAGCTAATA 477  
 QY 21 AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro 40  
 DB 478 GACAAACCCCACTTCCGAAAGAGGTGATGATCATACTCAAGTCCAGAGCAACACGCG 537  
 QY 41 ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60  
 DB 538 GTCAGAACCCCGCGCAATTTCTACTGTGGCAGAGACCTTGCAGCTGAATCGCCGACA 597  
 QY 61 GlyAlaGlyLysThrMetLysSerValValSerAlaAlaValGlnGlnLysLysArgMet 80  
 DB 598 GGAGCTGCCAAGACAATGAAAGCGTGATATCCGCTGTGTGCAGCAAAAGAGCGCATG 657  
 QY 81 SerAlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSerArgPro 100  
 DB 658 TCTGAGCAGCAGCGCCCTCCCTCCAGCAGACATGCGGAGTACTGCTCTTCGCGTCCC 717  
 QY 101 AlaAlaTrpAlaHisProProGlnAlaProLeuValGluLysAsnValTyrLysThr 120  
 DB 718 GCTGATCGGCACATCCACCTCCACAGGCTCTCTTGTGAGAGAAATGTATCAAGACT 777  
 QY 121 ProGlnGluGluProValTyrIleSerProGlnProArgSerLeuLysGluAsnLeuSer 140  
 DB 778 CCACAAGAGAGCCCGGTATACATATCACACAGCCTCGCAGTCTTTAAGGAAAAATCTAAGC 837  
 QY 141 ProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuAspPheThrProLeuThr 160  
 DB 838 CCTATGACGCCAGGAAACCTACTGACGCTGATGCAATCTCGGANTCACACCTCTCACC 897  
 QY 161 GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsnLeuAlaAlaTrpProThr 180  
 DB 898 GAAACCCGTGGCAAGGACAAGTACCATTTTCCGGACAAATCTGGCAGCCTGGCCACA 957  
 QY 181 ProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThr 200  
 DB 958 CCACACACTAAAGGGAATGTAATAATCATGTGCTAATGATATATCGCGCGCTCGAATCACT 1017  
 QY 201 ProAspAspLeuGluAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGlu 220  
 DB 1018 CCAGATGATCTAGAAGATCAGCTCCCAAAACAAACGTTTCGATGTAAGCATTCGAG 1077  
 QY 221 ThrIleAsnIleSerLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThr 240

DB 1078 ACTATCATATTTTGGTGACACCTTGGACTGCTCCAGATCGATGGACACCGCATAG 1137  
 QY 241 ProLeuAsnLysThrThrIleValHisAlaThrHisThrArgAlaLeuAlaCysIle 260  
 DB 1138 CCCCTAAATAAGACAACAACCATTTGTGCATGCCACGACACACAGAGCTCTGGCCCTGTATT 1197  
 QY 261 HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280  
 DB 1198 CATGAGGAGGAGGAGGACCAAGTCCACTAGGACGCCCAAGAGAGCCCATACACACCTG 1257  
 QY 281 LysArgAspIleLysLeuValGlySerProLeuArgLysTyrSerGluSerMetLysAsp 300  
 DB 1258 AAGAGGACATTAAGTTGGTGGGTTCAACCTTACGAAAGTATTCGAGTCCATGAAGAT 1317  
 QY 301 LeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsn 320  
 DB 1318 TTGTCACTTCTATCGCCACAAACTAAGTATGCAATTCGAAGGTCTATGCCTAATCTAAT 1377  
 QY 321 GluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlnGlnIleGlnIle 340  
 DB 1378 GAATGAAAAATCCGCTCGATCGAACAGAAATAGATATTACCAGAGCAGCAGATCCAGATA 1437  
 QY 341 LysAlaLysAspLeuAsnSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGlu 360  
 DB 1438 AAGCCAAAGACTTGAATAGCTCTCTAGTAGCAGGCTAGTTTGGCCGCGCAGCAGAG 1497  
 QY 361 PheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuIleGluVal 380  
 DB 1498 TTTCTATTCAACCAACAGTGCAGTGCATCTCGCTCAGTCCAGTCGTTTTTAATCTCCATGAAGTA 1557  
 QY 381 GlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGlu 400  
 DB 1558 GGTCCGAAGTCCGTGAAGGAGTCCGCTGAGGAATCTCTCAAGCGCGCTCTCATGAG 1617  
 QY 401 LeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGluThrValAlaIle 420  
 DB 1618 TTGAGTTTTTTCGGATGCACCTAGCAACGAATCATTTGATCCGCAATGAATGTAGCCATT 1677  
 QY 421 SerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAla 440  
 DB 1678 TCCCTCTCTAAAAAGCAACGGGTGGAGACACTACTCTGCGCCAGAAAGTGCAGCGCGCA 1737  
 QY 441 AsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPhe 460  
 DB 1738 AATGATCTGCAAGAGCAGTAGTCCCGCTGCGCACACGCCCAATCCAAAGAGTTT 1797  
 QY 461 LysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAsp 480  
 DB 1798 AAGCTAGCACAAACCATGTCTGATGAAGAGCCGCCACACACCAAGAGTCAAGGAC 1857  
 QY 481 ThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIle 500  
 DB 1858 ACTAGCATTCAGCCTTCCGTCAAGCTCTATGACTCGGAGCTGTATATGACAGCGTGCATC 1917  
 QY 501 AsnProAspProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyr 520  
 DB 1918 AACCCGATTCATTTGACGCAACTACGCAATTTGATTCCTTTCTGGCATCTACCAATGAT 1977  
 QY 521 LeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLysLysThrLeuAsnAlaLeu 540  
 DB 1978 TTGGATGAACAGGCTGTGATCGTCAATCAAGTGCATCTTTAAAGAGTGGTAAATGCCCTT 2037  
 QY 541 ValSerIleProAlaAspLeuAspAlaAspLeuAsnLysIleAspValGlyLysLeu 560  
 DB 2038 GTCTCCATACCCGCTGACCTGAGCGCAGATTTAAATAACAAATAAGAGCTTTGGTAAAGCTG 2097  
 QY 561 PheAsnGluValArgAsnLysGluLeuValAlaProThrLysGluGlnSerMet 580  
 DB 2098 TTTAACAGGTCGCGCAACAAAGAGCTCGTGGTGGCTCCCAACAGAGAGGAGCGATGTATG 2157  
 QY 581 AsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePhe 600

Db 2158 AACTACCTAACGAATAACCGCTGGAGACGCTTCGTAAGCGCGCTGTGGAGCTCTTCTTC 2217  
 QY 601 SerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeu 620  
 Db 2218 AGTGAGCAGATGCGCTTCGCTGCTCCAAAGTGGCGGTATATGTCAACAAAGCAAGCTCTG 2277  
 QY 621 ArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGlu 640  
 Db 2278 CGCATCCGTAGCATCGTATCTTCACCTAGACGTAGTATGCAACGCACCATACTTGAG 2337  
 QY 641 LeuLeuLeuCysPheAsnProLeuTyrPheLeuArgLeuGlyLeuGluValValPheGlyGlu 660  
 Db 2338 CTGCTGCTTTGCTTCATCCCTTTGGCTCGCGCTTGGACTGGAGTGGTCTTTGGCGAG 2397  
 QY 661 LysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArg 680  
 Db 2398 AAGATCCAGATGCGATCTAATCAGACATTTAGGCGCTCAGCACCTTTATCCCAATCGC 2457  
 QY 681 LeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu 700  
 Db 2458 TTGTTCCGGAAATAAGTGTGAGGAGCAGAGGTACAGCAAGCATACACACTCACCGAAGAG 2517  
 QY 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuProPheLeu 720  
 Db 2518 TAGCGGAGACCAATTAAGAAGCACTCATTCAGAAAATCCCTCTTTTGGCTGCTTTCCCTC 2577  
 QY 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740  
 Db 2578 GATCAAGCTTAGCAGAAAGCCATCGTCAAGCAATCCCTGTTGTTGTTTAAAGTCG 2637  
 QY 741 ProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIle 760  
 Db 2638 CCACATAAAGAGACCAAGGATATTCGTCGCTTCGTCGAGCTGCTCGCAACATT 2697  
 QY 761 GlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780  
 Db 2698 GGTGATATTACGGCGGAACCTCGTCGCTGGGTACGTTCTACAGCACCGCCAAACATT 2757  
 QY 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800  
 Db 2758 TTGACAGAGTTCGATATGCTTCAACAACTTGGCTGTGACCTAAGAGATGGCGTGAGA 2817  
 QY 801 LeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820  
 Db 2818 CTAACCCGAGTCATCGAGGTAAATTTTACTACGCGATCTAACCCTGCGCAATTAAGGGTG 2877  
 QY 821 ProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840  
 Db 2878 CCAGCCATCTCTCGCCTTCAGCGGATCTTCAATGTAAAGTAGCTCTGGCGCACTTGGT 2937  
 QY 841 GluAlaAsnPheGlnLeuGlyGlyValAspIleAlaAlaGlnAspIleValAspGlyHisArg 860  
 Db 2938 GAAGCCAACTTCCAGTAGCGCGGACATCGCGCCCAAGACATCGTTGACGACATCGT 2997  
 QY 861 GluLysThrLeuSerLeuLeuTyrPheGlnLeuIleTyrLysPheArgSerProLysPheHis 880  
 Db 2998 GAGAAAGCGCTTCCCTGCTCTGGCAACTTATTTACAAATTCGCTCGCCCAAGTTTCAT 3057  
 QY 881 AlaAlaAlaThrValLeuGlnLysTyrTyrArgArgHisTyrPheHisValValIleGln 900  
 Db 3058 GCGCGCGCACCGGTGTCCAGAAATGGTGGCGCGCTACCTGGCTGCACCGTTGTATCCAG 3117  
 QY 901 ArgArgIleArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAla 920  
 Db 3118 CGTCGATTCGCCACAAAGAGCTTATGGCTCGCCACCGGGCCGCTACTGTTCATTCAGGCC 3177  
 QY 921 ValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGln 940  
 Db 3178 GTGTTCCGTGGCCACCAGATGAGAAAGTACGTGAAGTTGTTTAAAGCGGAACCACTCAG 3237  
 QY 941 AlaAlaIleIleLeuGlnLysPheThrArgArgTyrIleuAlaGlnLysGlnLeuTyrGln 960  
 Db 3238 GCCGCAATAATCTGCAAAAAGTTTCAACCCGTCGGTATTTGGCCCCAGAGCAAGCTATATCAG 3297

QY 961 SerTyrHisSerIleIleThrIleGlnArgTyrTyrArgAlaGlnGlnLeuGlyArgGln 980  
 Db 3298 AGCTATATCAAGTATATATCATCATCCAGCGCTGTGTGGCGAGCCCAACAACACTGGGAGGCAG 3357  
 QY 981 HisArgGlnArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTyrArg 1000  
 Db 3358 CACCGCCAGCGGTTTGTGGAGCTCCGAGAGGCTCGCATCTTTCTTCAGCGGATCTGCGG 3417  
 QY 1001 ArgArgLeuPheAlaLysLysLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSer 1020  
 Db 3418 CGACGACTCTTTGCCAAAAAATAATTTGGCGCGCGGAAACACAGCACTTCAGCGCATCG 3477  
 QY 1021 GlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTyrArgThrTyrGlnLeuGly 1040  
 Db 3478 CAAAAACAACAGCAGCTGCTAGTATATTCAAATGCATGGCGAAGCTATCATCTGGCG 3537  
 QY 1041 ArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArg 1060  
 Db 3538 AGAATTCAGCGACAGCAGCTTCCTGCGCCAGAGGAGCTCATCATCTGTTGTTTCAGCGCAGG 3597  
 QY 1061 MetArgSerLysTyrSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAla 1080  
 Db 3598 ATCGGAAGCAAGTGGAGTATGCTGGAGCAGCGCAAGAGTTCAGCAACTAAAGCGTGCA 3657  
 QY 1081 AlaIleAsnIleGlnGlnArgTyrArgAlaLysLeuSerMetArgLysCysAsnAlaAsp 1100  
 Db 3658 GCTATAAATATCAACAACGCTGCGCGAGCAAGCTTTCANTGAGAAGTGCACGCTGAT 3717  
 QY 1101 TyrLeuAlaLeuArgSerValLeuLysValGlnAlaTyrArgLysAlaThrIleGln 1120  
 Db 3718 TATTTGGCACTTCGTTCCAGCGTCTTAAAGTTTCAGGCTTACAGGAAAGCACAAATCCAG 3777  
 QY 1121 MetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGln 1140  
 Db 3778 ATGAGAATAGATCGTAAATCACTACTATTCCTCGGAAAAAATGTATCTGCTCGTGCACAG 3837  
 QY 1141 ArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsn 1160  
 Db 3838 CGACTCAGGGCCATCATGAAATGCGGCAACAGAGGGGAAATATCTGAGGCTCGCAAT 3897  
 QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnMetIleGlnAspArgAsn 1180  
 Db 3898 GCTTCAATACTAGTTCAAAAACGCTACCGCATCGTCAACAAATGATCCAGGATAGAAAT 3957  
 QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTyrArgAlaThrLeu 1200  
 Db 3958 GCATATTTAAGAACCCGCAATGTATCAATGTCCAGAGGCGCTGGAGGCCACTCTG 4017  
 QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
 Db 4018 CAGATCGCTCGAGAAAGGAAGAACTACCTTCATCTCAAAACGACAAACCAATTCAA 4077  
 QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys 1240  
 Db 4078 ATCAAGTTCGTCGCCAAGCGCAATGAAAGCAAGAGAGCGAGTTCCTTCAGCTGAAA 4137  
 QY 1241 LysValThrLeuValValGlnLysArgArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 Db 4138 AAGTTTACCTTGTAGTCCAGAAACGTGGCGGCGCTTGTCTGCAGATCGGAAGAGGCGC 4197  
 QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280  
 Db 4198 CAGGAGTACCTCATCTGCGCGAGGTTACCATAAAGCTGCAGCGCAGATTCATGTCTCAA 4257  
 QY 1281 LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu 1300  
 Db 4258 AAATCAATCGGTTCTATCGAGCAAGTACCCGCGCACCCAGGCTGCTGTGAGCTGCTG 4317  
 QY 1301 GlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeu 1320  
 Db 4318 CAGATGCATTTGGGTAAATCATCTCTTTAGGAAACGGGAGGAGGACAGTTTCTTACAAC 4377



XX  
PS Claim 1; SEQ ID NO 30122; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins (AB057737-  
CC AB072072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO.int at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 8330 BP; 2321 A; 2009 C; 1949 G; 2051 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 8330  
Score: 9386.50 Matches: 1855  
Percent Similarity: 96.32% Conservative: 4  
Best Local Similarity: 96.11% Mismatches: 2  
Query Match: 98.66% Indels: 69  
DB: 4 Gaps: 3

US-09-914-698-1 (1-1861) x AB111880 (1-8330)

QY 1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeu 20  
DB 1542 ATGGAGCTAGTGTGGAGCCCGCTCTTGTAGTGGCTGCGAGAGAGCGTGCAGCTAATA 1601  
QY 21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerAsnGlnPro 40  
DB 1602 GACAAACCGCAACTCCGAAAGAGGTGATGATCATCTCAAGTCCAAAGACCAACGCG 1661  
QY 41 Val----- 41  
DB 1662 GTCAAGGTGAGTCAAAATATGCAATTTGCCACCCCTTTTACTATACATACCCCTTTTGCAT 1721  
QY 42 -----LysAsnProArgLysPheProThrValGlyLysThrLe 54  
DB 1722 GTATCTCTTACCTTTCTACTGAGAACCGCGCAATTTCTTACTGTGCGCAAGACCCCT 1781  
QY 54 uGlnLeuLysSerProThrGlyAlaGlyLysThrMetLysSerValIsetalAlaVa 74  
DB 1782 GCAGCTGAAATCCCGACAGAGCTGGCAAGACAATGAAAGCGTGATCCGCTCTGT 1841  
QY 74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94  
DB 1842 GCAGCAAAAGAACGCGATGCTGCAGCAGCAGCGCTCCCTCCAAAGCAGACATGGCGAGT 1901  
QY 94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValGl 114  
DB 1902 GACTGCTCCTTCGCTCCGCTGCATGGGACATCCACCTCCACAGGCTCCTCTTGTGCA 1961  
QY 114 uLysAsnValTyLysThrProGlnGluProValTyIleSerProGlnProArgSe 134  
DB 1962 GAAGATGTATACAGACTCCCAAGAGAGCCGTTATACATATACCCACAGCCTCGCAG 2021  
QY 134 rLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154  
DB 2022 TCTTAAGGAAATCTTAAGCCCTTATGACGCGAGAAACCTTACTCGAGTGAATGACAAATCT 2081  
QY 154 uArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePheProAspAs 174  
DB 2082 GCGATTCACACCTCTCACCGAAACCGTGGCAAGGACAGCTACCATTTTCCCGACAA 2141  
QY 174 nLeuAlaAlaTrpProThrProThrLysGlyAsnValLysSerCysAlaAsnAspMe 194  
DB 2142 TCTGGCAGCTGGCCACACCAACACTTAAAGGGAATGTAATAATCATGTGCTTAATGATAT 2201  
QY 194 tArgProArgArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPh 214  
DB 2202 GCGGCCGCGCTCGAATCACTCCAGATGATCTAGAAGATCAGCTGCCACAAACAAACGTT 2261

QY 214 eAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAspCysSerArgI 234  
DB 2262 CGATGTAAGCATTCCGAGACTATCAATATTTCTGGACACCTTGGACTGCTCCAGAT 2321  
QY 234 eAspGlyGlnProHisThrProLeuAsnLysThrThrIleValHisAlaThrHisTh 254  
DB 2322 CGATGACAAACCGCATACGCCCCCTAATAAGACAACACCATTTGTCATGCCACGCACAC 2381  
QY 254 rArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProProArgThrProThrLy 274  
DB 2382 CAGAGCTCTGGCTGTATTTCATGAGGAGGAGGACCAAGTCCACCTAGGACGCCACGAA 2441  
QY 274 sSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTy 294  
DB 2442 GAGCGCATACACACCTGAGAGGGACATTAAATGTTGGTGGTTACCCCTTACCAAGTA 2501  
QY 294 rSerGluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyAlaIleGlnI 314  
DB 2502 TTCGAGTCCATGAAAGATTGTCACTTCTATCGCCACAACTAAGTATGCCATTCAAGG 2561  
QY 314 ySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyTrpGl 334  
DB 2562 GTCTATGCCCTAATCTAAATGAAATGAAATCCGCTCGATCGAAACAGATATATACCA 2621  
QY 334 nGluGlnGlnIleGlnLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSe 354  
DB 2622 GGAGCAGCAGATCCAGATTAAGGCCAAAGACTTGAATAGCTCTCTAGTAGCGAGCTAG 2681  
QY 354 rLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerAr 374  
DB 2682 TTTGGCGGCCAGCAGAGTTTCTATTCAACACACAGTGAGATCTCTCGCTCAGTCCAGTCG 2741  
QY 374 gPheAsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHi 394  
DB 2742 TTTTAATCTCCATGAAGTAGTGGAGTCCGTAAGGGAAGTCCGCTGAAAGATCTCTCA 2801  
QY 394 sLysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGlnSerLeuTyAr 414  
DB 2802 CAAAGCCCGCTCTCATGAGTTGAGTTTTTCGGATGCACCTTAGCAAGATCATTTGACCG 2861  
QY 414 gAsnGluThrValAlaIleSerProPolyLysGlnArgValGluAspThrThrLeuPr 434  
DB 2862 CAATGAAACTGTAGCCATTTCCCTCTTAAAGCAACGGTTGAGGACACACTACTCTGCC 2921  
QY 434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerSerAlaHiAlaTrpProHi 454  
DB 2922 CAGAAGTGCAGCGCGCGCAATGTCATCTCAAGAACAGTAGTGCCTCCACGCTGGCCACA 2981  
QY 454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
DB 2982 CGCCCAATCCAGAGTTTAAGTAGCAACCAACATGTCTAGTAGAAGAGAGGCCGCCAC 3041  
QY 474 rProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyAspSerGluLe 494  
DB 3042 ACCACGAAAGTCAGGACACTAGCATTGAGCTTCCGCTCCGCTCTATGACTCGGAGCT 3101  
QY 494 uTyThrMetGlnThrCysIleAsnProAspProPheAlaAlaThrThrIleAspProPh 514  
DB 3102 GTATATGACAGCTGCATCAACCCGATCCATTTTCAGCAACTAGACAAATGATTCATT 3161  
QY 514 eLeuAlaSerThrMetTyLeuAspGluGlnAlaValAspArgHisGlnAlaAspPheLy 534  
DB 3162 TCTGGCACTCATCATGTATTTCGATGAACAGGCTGTGGATCGTCATCAAGTACTTTAA 3221  
QY 534 sLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsnLy 554  
DB 3222 AAAGTGGTTAAATGCCCTTGTCTCCATACCCGCTGACCTGGACGACAGATTTAAATAACAA 3281  
QY 554 sIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeuValValAlaProTh 574  
DB 3282 AATAGACGTTGTAGCTGTTTTAAAGGTGCGCAACAAAGAGCTCGTGTGCTGCCAC 3341



Qy 574 rLysGluGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAl 594  
Db 3342 CAAGAGAGAGAGCTATAGAACTACCTAACGAATACCGCTGGAGAGCGCTTCGTAAAGC 3401  
Qy 594 aAlaValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTy 614  
Db 3402 GGCTGTGGAGCTCTTCTCACTGAGCAGATCGCTGCCATCTCCAAAGTGGCGGTATA 3461  
Qy 614 rValAsnLysGlnAlaLeuArgLysLeuArgSerAspArgAsnLeuHisLeuAspValValMe 634  
Db 3462 TGTCAACAAGCAAGCTCTGCGCATCCGTAGAGATCGTAATCTCACCTAGACGTAGTTAT 3521  
Qy 634 tGlnArgThrLysLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLe 654  
Db 3522 GCAACGACCAATCTTGAGCTGCTTGTGCTTCAATCCCTTTGGCTGGCGCTTGGACT 3581  
Qy 654 uGluValValPheGlyGluLysLysLeuGlnMetGlnSerAsnArgAspLysValGlyLeuSe 674  
Db 3582 GGAAGTGGTCTTTGGCGAGAAAGATCCAGATCGACTCAATCGAGACATTTGTAGGCTCAG 3641  
Qy 674 rThrPheLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAl 694  
Db 3642 CACCTTTATCTCAATCGCTTGTTCGGGAATAAGTGTGAGGAGCAGAGGTACAGCAAGGC 3701  
Qy 694 aTyrThrLeuThrGluGluTyrAlaGluThrLysLysLysHisSerLeuGlnLysLysLe 714  
Db 3702 ATACACACTACCGAAGAGTAGCGGAGACCATTAAGAACGACCTCATTTGCAGAAATCCT 3761  
Qy 714 uPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgLysValLysHisAsnProCy 734  
Db 3762 CTTTTTGTCTTTTCTCGATCAAGCTAAGCAAGAGCGATCGTCAAGCAACATCCCTG 3821  
Qy 734 sLeuPheValLysLysSerProHisLysGluThrLysAspLysLeuLeuArgPheSerSe 754  
Db 3822 TTTCTTTGTTAAAGTGCACACATAAGAGACCAAGGATATTCCTGCTGCGCTTCTCGTC 3881  
Qy 754 rGluLeuLeuAlaAsnLysLysGlyAspLysLeuArgLysLeuArgLysLysValLe 774  
Db 3882 GGAGCTGCTGCCAACATTTGTTGATTTACCGGGAATCTCGTCCCTGGGCTTACGTTCT 3941  
Qy 774 uGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAs 794  
Db 3942 ACAGCAGCGCAACATTTTGGACGAGTTCGATATTCCTTCAACACTTGGCTGTGA 4001  
Qy 794 pLeuArgAspGlyValArgLeuThrArgValValGluValLysLeuLeuArgAspLysLe 814  
Db 4002 CCTAAGAGATGGCTGAGACTAAACCGAGTCTATGAGGTAATTTTACTACGCGATGATCT 4061  
Qy 814 uThrArgGlnLeuArgValProAlaLysSerArgLeuGlnArgLysPheAsnValLysLe 834  
Db 4062 AACCGCCAGTTAAGGTGCCAGCATCTCTCGCTTCAGCGGATCTTCAATGTAAAGCT 4121  
Qy 834 uAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspLysLeuAlaGlnAs 854  
Db 4122 AGCTCTGGCGCACTTGTGTGAAGCAACTCTCCAGCTAGCGGCGACATCGCGCCCAAGA 4181  
Qy 854 pLysValAspGlyHisArgGluLysThrLysSerLeuLeuThrGlnLeuLysLysPh 874  
Db 4182 CATCGTTACGGACATCGTGTGAGAGACGCTTTCCTGCTCTGGCAACTTATTTTCAAAAT 4241  
Qy 874 eArgSerProLysPheHisAlaAlaAlaThrValLeuGlnLysTyrTrpArgArgHisTr 894  
Db 4242 CCGCTCGCCCAAGTTTCAAGCGCGGCCACCGTGTCTCAGAAATGGTGGCGGCTCACTG 4301  
Qy 894 pLeuHisValValLysLeuGlnArgArgLysGlnLysGluLeuMetArgArgHisArgAl 914  
Db 4302 GCTGCGCTTGTATCCAGCGTCCGATTCGCCACAAAGAGCTTATGCTCGCCACCGGCG 4361  
Qy 914 aAlaThrValLysLeuGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPh 934  
Db 4362 CGCTACTGTCATTCAGGCGGCTGTTCCGTGGCCACCGACATGAGAAAGTACGTGAAGTTGT 4421  
Qy 934 eLysThrGluArgThrGlnAlaAlaAlaLysLysPheThrArgArgTyrLeuAl 954

Db 4422 TAAGCGGAAGCACTCAGCGCGCAATAATTCGCAAAAGTTCCACCGTCGGTATTGGC 4481  
Qy 954 aGlnLysGlnLeuTyrGlnSerTyrHisSerLysLysLeuGlnArgTyrTrpArgAl 974  
Db 4482 CCAGAGCAGCTATATCAGAGCTATCAAGTATATCACCATCCAGCGCTGGTGGCAGC 4541  
Qy 974 aGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAlaLysPh 994  
Db 4542 CCAACAATCGGAAGGAGCACCAGCGGCTTGTGGAGTCCGAGAGGCTCGCATCTT 4601  
Qy 994 eLeuGlnArgLysLeuPheArgLeuPheAlaLysLysLeuLeuAlaAlaGluTh 1014  
Db 4602 TCTTCAGCGGATCTGGCGGCGACCTCTTTCGCAAAAACATTTGGCGCGCGGAAC 4661  
Qy 1014 rAlaArgLeuGlnArgSerGlnLysGlnAlaAlaAspSerTyrLysGlnMetGlnTr 1034  
Db 4662 AGCCAGACTTCAGCGATCGCAAAAACACAGCGAGCTGTAGTTATATTCAAATGCAATG 4721  
Qy 1034 pArgThrTyrGlnLeuGlyArgLysGlnArgHisGluPheLeuArgGlnArgAspLysL 1054  
Db 4722 GCGAAGCTATCAGCTGGCGAGATTCAGCGACACAGCGAGTCTCTGCGCAGAGGACCTCAT 4781  
Qy 1054 eMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPh 1074  
Db 4782 CATCTTTGTTTCAGCGCAGGATCGAAGCAAGTGGAGTATGCTGGAGCAGCGCAGGAGTT 4841  
Qy 1074 eGlnGlnLysArgAlaAlaLysLysGlnArgLysGlnArgTrpArgAlaLysLeuSerMe 1094  
Db 4842 CCAGCAACTAAAGCGTCAGCTATATAATATCCAAACACGCTGGCGAGCGAAGCTTCAAT 4901  
Qy 1094 tArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLysLysValGlnAlaTy 1114  
Db 4902 GAGAAAGTGCACGCTGATTTTGGCACTTCGTTTCAGCGCTTCTTAAAGTTTCAGGCTTA 4961  
Qy 1114 rArgLysAlaThrLysGlnMetArgLysLysAspArgAsnHisTyrTyrSerLeuArgLysAs 1134  
Db 4962 CAGGAAGCCACATCCAGATGAGATAGATCGTAATCACTACTTATTCCTCGCGAAGAAA 5021  
Qy 1134 nValLysCysLeuGlnArgLeuArgAlaLysMetLysMetArgGluGlnArgGluAs 1154  
Db 5022 TGTATCTGCTGCAACAGCGACTGAGCGCCATCATGAAATGCGCGAAGCAAGAGGAAAA 5081  
Qy 1154 nTyrLeuArgLeuArgAsnAlaSerLysLeuValGlnLysArgTyrArgMetArgGlnGl 1174  
Db 5082 TTATCTGAGGCTCGAATGCTTCAATACTAGTTTCAAAAACGCTACCGCATCGCGTCAACA 5141  
Qy 1174 nMetLysGlnAspArgAsnAlaTyrLeuArgThrArgLysCysLysLysValGlnArg 1194  
Db 5142 AATGATCCAGATAGAAATGCATATTTAAGAACCCGCAANTGATCATCATGTCACAG 5201  
Qy 1194 gArgTrpArgAlaThrLeuGlnMetArgArgLysLysAsnTyrLeuHisLeuGlnTh 1214  
Db 5202 CGCGTGGAGAGCCACTCTGCGATCGCGTCGAGAAAGGAAGAACTACCTTCATCTCCAAAC 5261  
Qy 1214 rThrThrLysArgLysGlnLysPheArgAlaLysArgGluMetLysLysGlnArgAl 1234  
Db 5262 GACAAACCAACGAATTCAAATCAAGTTCGCGCCAGCGGAAATGAAAAGCAAGAGAGC 5321  
Qy 1234 aGluPheLeuGlnLysLysValThrLeuValValGlnLysArgArgAlaLeuLe 1254  
Db 5322 CGAGTTCTTCAGCTGAAAAGGTTACCTTGTAGTCCAGAAACGTCGCGCGGCTTGTCT 5381  
Qy 1254 uGlnMetArgLysGluArgGlnLysLeuHisLeuArgGluValThrLysLysLeuGl 1274  
Db 5382 GCAGATCGAAAGGAGCGCCAGGAGTACCTACATCTGCGGAGGTACCATAAAGCTGCA 5441  
Qy 1274 nArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGl 1294  
Db 5442 GCGAGATTCATCTCAAAATCAATCGCGGTTTCATCGAGCCCAAGTACCGCGGCCCA 5501  
Qy 1294 nAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluAr 1314

Db 5502 GGCTGCTGTGAGCTGCTGCAGATGCTTGGCGCTAATCATCTCTCTTAGGAAACGGGAGAG 5561  
QY 1314 gAenSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaAr 1334  
Db 5562 GAACAGTTTCTTCAACTGGTCAAGCAGCAATAACACTGCAGCGACGGTACCAGCTCG 5621  
QY 1334 gLeuAenMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIl 1354  
Db 5622 TCTGAATATGATCAACAGCTTGAAGAGTTACGCCAGCTGAACAGCAGCTATTACCAT 5681  
QY 1354 eGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGl 1374  
Db 5682 TCAACCCCGATATAGAGCCCAAGAGGCAATGCAAAAGCAGGTGCTTGTACCAAAAGCA 5741  
QY 1374 nArgGluAlaIleLysValGlnArgArgTyrArgGlyAsnLeuGlnLeuMetArgLysGl 1394  
Db 5742 AAGGGAAGCCATTATCAAGTGCACACGACATACCGCGCAATCTCGAGATGAGGAAGCA 5801  
QY 1394 nIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrPrpArgSe 1414  
Db 5802 GATTGAAGTCTACCAAAACAGCGCCAGGCAGTCATCGCTTCGAGAAATGGTGGCGCAG 5861  
QY 1414 rIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSe 1434  
Db 5862 TATACCGGCATCGCGCTGTGCAAAAGCGGGCTACCGAAGGATTTCGACTCAGTTTCAATTGAG 5921  
QY 1434 rIleGlnArgLysTyrArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSe 1454  
Db 5922 CATTCACGCAAGTGGCGGCCACAGTCGAGGCTCGTCCCAACGGGAGATCTTCTTAAG 5981  
QY 1454 rThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGl 1474  
Db 5982 CACCATCCGCAAAAGTCGACTTATGACGAGCAATTTACTGATGCGTCA 6041  
QY 1474 nGlnArgArgGluPheGlnMetLysArgArgAlaAlaValValIleGlnArgArgPheAr 1494  
Db 6042 GCAGCGCAGGGAATTCAGATGAAAGGAAAGGCTCGCTAGTATCCAGCGCGGTTTCG 6101  
QY 1494 gAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIl 1514  
Db 6102 TGCTCGCTGTGCTATGCTTAAGCGGAGACAGGATTAACCAATTAATCCATCTCTGTGAT 6161  
QY 1514 eLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheVa 1534  
Db 6162 CCTGGTGCAGCGCAAAATTCGTCCTAATCGCAGCATGAAGCAGCGCACCGAGAAATTGT 6221  
QY 1534 lGlnLeuArgThrIleAlaValHisLeuGlnLysPheArgGlyLysArgLeuMetIl 1554  
Db 6222 CCAGCTGCGTACTATTGCGATGCTCATTTGCAACAAAGATTCCGTTGGCAAGCTCTAATGAT 6281  
QY 1554 eGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAl 1574  
Db 6282 TGAGCAGCGTAATTGTTTCCAACCTGCTCGCTGCTTCATGCCGGGTTTCAGGCGAGCTGC 6341  
QY 1574 aArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLe 1594  
Db 6342 CCGCGGCTTTATGGCTCGCAAAACATTCAGGCGCTGATGACACCCGAGATGATGACCT 6401  
QY 1594 urIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrIleuIleAr 1614  
Db 6402 CATCCGCCAAGACGCGCCCAAGGTTATACAGCGTTACTGCGGGGCTATCTAATCCG 6461  
QY 1614 gArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGl 1634  
Db 6462 ACGACCTCAGAGCACCGGGTCTCTTGATATCCGAAGCGTATTGCTCAGCTGCGGCA 6521  
QY 1634 nGluAlaLysAlaValAsnSerValArgCysLysValGlnGlnAlaValArgPheLeuAr 1654  
Db 6522 AGAGGCAAAAGCGGTAAATCTCTGCGCTGCAAAAGTCCAGGAGCGGTTCGTTCTCTCG 6581  
QY 1654 gGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu- 1669  
Db 6582 CGACGCTTTATCGCATCAGATCGGTTAGCAGTGCCTAAGTTCGATTGGGTAAGTGATCCAG 6641

QY 1670 -----AspArgLeuSerAr 1674  
Db 6642 CTACGTTAGTCTCTATTATATAACTAATCCTGCTTAAATAATCCCGCAGATCGTCTTCGCG 6701  
QY 1674 gThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIl 1694  
Db 6702 CACTGTGCCACACCTGCTCATGTGTGGTTCGGAGTTTCATGTCACGTTTGTCTATGGCAT 6761  
QY 1694 eMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgIl 1714  
Db 6762 CATGGCTCAGGCCATTCGATCAGAGGTGATAGCAGCTTATAGACGCTGCACGCCGAT 6821  
QY 1714 eIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGluGlyLysLe 1734  
Db 6822 CATCTTAATTTGGCCCGCTACATAGCACCCAGCGTGAACACGTTCCAGGAGGCGGTTT 6881  
QY 1734 uValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluIlePheAsnTh 1754  
Db 6882 GGTCAACATTGCCAGATGTTATTGGCTGGTGCAGACAAAGACAGTGAGATATTCAACAC 6941  
QY 1754 rIleuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLys----- 1770  
Db 6942 TTTGTGCACCCCTCATTTGGGTATTTCGCCCACTGTCTTAAAGCGAAAGGTAGCTAAATA 7001  
QY 1771 -----IleIleH 1773  
Db 7002 GCGAGGAGATTAGTTTATAGCGAGTTTAAATCATATTCTTAATGCTTCCAGATCATTC 7061  
QY 1773 iAspTyrMetThrAsnProGluAlaIleTyrMetValArgGluThrLysLysLeuValA 1793  
Db 7062 ACGACTATATGACCAACCCAGAGGCCATTTACATGTTGCGGCAAACTAAGAAGCTCGTGG 7121  
QY 1793 lAArgLysGlnLysMetLysGlnAsnAlaArgLysProProMetThrSerGlyArgT 1813  
Db 7122 CCCGCAAGGAAAGATGAAGCAAAATGCACGCAAGCCGCGCAATGACAAAGTGCAGCT 7181  
QY 1813 rLysSerGlnLysIleAsnPheThrProCysSerLeuProSerLeuGluProAspPheG 1833  
Db 7182 ATAGAGCCCAAGATAACTTCACGCCGTGTTCTGCGCCAGCTGGAGCGGACTTCG 7241  
QY 1833 lYIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAspThrIleL 1853  
Db 7242 GAATCATCGCTACAGTCCCTACAGCTTTATCTCGTCCGTTTACGCTTTCGATACGATT 7301  
QY 1853 euCysLysLeuGlnIleAspMetPhe 1861  
Db 7302 TGTGCAAGCTGCAGATCGACATGTTT 7327  
RESULT 4  
AAH14237  
ID AAH14237 standard; cDNA; 2453 BP.  
XX  
AC AAH14237;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:11532.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX OS Homo sapiens.  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.





```
QY      890  -----Tr 890
Db      1443  TGACGCAAACTGGTTCAGTGGTATTAATAATCATCATCTGAATGATGACAGTTCCTG 1502
QY      890  pArgArgHisTrpLeuHisValVal
Db      1503  GATATGTCACCTTAAAGCATTTGATGAAATACCTTTCAGAGCTATACAAAGAGCTCCTA 1562
QY      899  -----IleGlnArgArgIleArgHis
Db      1563  GAAATGAAAGAAAGAAATTTT-CACCTTGGTTAGTCTGCGAGTTAGAGACCTTGGTGAAT 1621
QY      905  -----
Db      1622  ACCTGCTATGATTAATCATTCAGATATGTCATCAATTCAGATCAAAAGGTGGTTAT 1681
QY      906  -----LysGluLeuMetMetArgHisArgAlaAl
Db      1682  TACCTATTTTCTCATTTCTTTGTGCAAGCTTTTGGATCTTCGTAAGAAATAAGAGCTGC 1741
QY      915  aThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrrValLysLeuPheLy 935
Db      1742  TCGACTCATACAACACATCGGAGAAATATAAACTAAACACAGATCTCAACGCCCATCA 1801
QY      935  sThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyrrLeuAlaG1 955
Db      1802  G---GAGAGAGAGAAGCTCGAAGATTTATCAATTCGCTGTAATCAATTTCTAGCAAA 1858
QY      955  nLysGlnLeuTyrrGlnSerTyrrHisSerIleIleThrIleGlnArgTrpArgAlaG1 975
Db      1859  ACAAGATTGAGAAAGAGAGTAAATGCGCACTCGTCAATTCAGAAATATTGG----- 1910
QY      975  nGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaIlePheLe 995
Db      1910  -----
QY      995  uGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015
Db      1911  -----CGAAGATCTTAGCAGAGAGAAATTTAATGTTAAAGAAAGGA 1954
QY      1015  aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAlaSerTyrrIleGlnMetGlnTrpAr 1035
Db      1955  AAAGCTGGAAAGATTCAAAATAAA-----GCAGCATCACTTATTCAGCGATATTGGAG 2008
QY      1035  gThrTyrrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMe 1055
Db      2009  AGATATTCC-----ACTAGACAAAGATTTCGAAATGGAATTAATTATTCAT 2056
QY      1055  tPheValGlnArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGluPheG1 1075
Db      2057  CATCTGCAATCTAGGATAAGA-----ATGATAATTGCTGTTACATCTTATAA 2104
QY      1075  nGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetAr 1095
Db      2105  ACCGATCTTTGGGCTACACTTCAATTCAGAGCATTCGCGTCTTATTAAAGAGAA 2164
QY      1095  gLysCysAsnAlaAspTyrrLeuAlaLeuArgSerValLeuLysValGlnAla---Ty 1114
Db      2165  ACAGATCAACAAGATATGAAATGCTGTAATCAATCACTTATTAATCAATCTATGTT 2224
QY      1114  rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrrSerLeuArgLysAs 1134
Db      2225  CAGAAAA-----TGGAAGCAACGTAATA-- 2246
QY      1134  nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgG1 1153
Db      2247  -----ATGCAATCAACAAGTAAAGCTACAGTAATTAATTCGAAGAGCTTTTGA 2296
QY      1153  uAsnTyrrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169
Db      2297  ATGGCATTAAAGAAACAAAGCAAGTAAGAGAAATTTCTGCTATTATCATACATCATGGTA 2356
```

```
QY      1169  rArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrrLeuArgThrArgLysCysI1 1189
Db      2357  TAGAATGCATAAAGATTA-----CGGAAGTATATTATATAGATCTTGTGT 2404
QY      1189  eIleAsnValGlnArgArgTrpArg 1197
Db      2405  TGTATCATTCAGAAAAAGATTCCGG 2429
RESULT 5
AAH14236
ID   AAH14236 standard; cDNA; 3887 BP.
XX
AC   AAH14236;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11530.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PE   28-JUL-2000; 2000EP-00116126.
XX
PF   29-JUL-1999; 99JP-00248036.
XX
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR   WFI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 11530; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesizing 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesizing polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 3887 BP; 1436 A; 652 C; 760 G; 1039 T; 0 U; 0 Other;
```

Alignment Scores:			
Pred. No.:	2,02e-41	Length:	3887
Score:	707.50	Matches:	276
Percent Similarity:	40.70%	Conservative:	225
Best Local Similarity:	22.42%	Mismatches:	412
Query Match:	7.44%	Indels:	319
DB:	4	Gaps:	40
US-09-914-698-1 (1-1861) x AMH14236 (1-3887)			
Qy	868	TpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln	887
Db	359	TGGATGATGAAGAAAGGATCGA-----GAGATGCACAGGGCTGTCTTTCATCCAG	412
Qy	888	LysTrpTrpArgArgHisTrpLeuHis-----Val	897
Db	413	TCTACTTTTCAGATGACAGATTACATATGAGATATCAGGCTTTGAAACAGGCCTCCGTT	472
Qy	898	ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet	909
Db	473	GTGATCCCAACAGCAATACCAAGCAAAATAGAGCTGCAAAACTGCAGAGCAGCATATATCTC	532
Qy	910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929
Db	533	AGACAAAGACACTCTGCTGTGATCCTTCAGGCTGCATTTCAGGGGTATGAAACTAGAAAGA	592
Qy	930	TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThr	949
Db	593	CATTGTGAAGAGTATGCATTCC-----TCTGCAACCTTTATTTCAGAGTAGGTTT	640
Qy	950	ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969
Db	641	AGATCAITACTGGTGAGGAGAGATTCATTTCCTCCCAAAAGCTACTATTTTGTGTCAG	700
Qy	970	ArgTrpTrpArgAlaGlnLysGlnLeuGlyArgGlnHisArgGlnArgPheValGlnLeuArg	989
Db	701	AGGAAATATCAGACCACCATTTTGTGCCAAACATAAATTTGTACCAATTTCTGCACCTAAGA	760
Qy	990	GluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeu---	1008
Db	761	AAGCGACCATTTACATACAGTCATCTTACAGACACTCATGTTAAAGAAAGAGTTACAA	820
Qy	1009	-----LeuAlaAlaAlaGlnThrAlaArgLeuGlnArgSerGlnLysGln---	1023
Db	821	GAATGCAAAAGGGCTGCAGT--TCTCATTCAGGCTACTTTCAGATGCACAGCAAAAAGAAA	879
Qy	1024	-----GlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThr	1036
Db	880	TATATTACATTTTCAGACTTGGAAACATGCTTCAATTTCTAAATTCAGCAACATTTATCGAACA	939
Qy	1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspIleMetPhe	1056
Db	940	TATAGAGCTGCAAAATTCGAAAGAGAAATTTATATACAGCAATGCGATTTCTGCTGTGTT	999
Qy	1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGlu	1073
Db	1000	ATTTCAGGCTGCATATAAAGGAATGAAGCAAGCAACTTTTAAAGGAAACACAAAGCT	1059
Qy	1074	-----PheGlnGlnLeu	1077
Db	1060	TCTATTGTAATACAGGCACCTACAGATGTATAGCAGTATTGTTTCTPACCAAAAGCTT	1119
Qy	1078	LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097
Db	1120	CAGTGGCTTACAAAATCATCAAGAAAAATATAGACCAATATAAAGAAACACAGAAA---	1176
Qy	1098	AsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----ValGlnAla	1113
Db	1177	-----GTATTTCAACAATGAACCTTAAGAAAAAGAGACTTGTGTTCAGGCA	1221
Qy	1114	---TyrArgLysAlaThrIleGlnMetArgIleAsp-----	1124
Db	1222	GGTTTCAGGACATGACATATAAAGAAACAGATTCAGAAACAGCAGCAGCCAGCTGCCATTATT	1281

Qy	1125	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1134
Db	1282	ATTTCAGAGCAATTTGTAAGCCCTTTAAATAAGGAAGCAATTAATCTCCCACTTAGACACACA	1341
Qy	1135	ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	1151
Db	1342	GTAGTTTCTATTCAAAGAAAGATACAGAAAACTAACTGCAGTCGCTTACCAACAGCACTTATT	1401
Qy	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
Db	1402	TGTATACAGTCTTATTACAGAGGCTTTAAAGTACCAAGAGGATATTCAAAATATGCACCGG	1461
Qy	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
Db	1462	GCTGCCACACTAATTCAGTCAATTCATCGAATGACAGGGCCAAAGTTGAT-----	1512
Qy	1181	AlaTyrLeuArgTrpArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
Db	1513	---TATGAACAAGAAAACTGCAATTTGTTGTTATACAGAATTTATTAGGTTGTATGTT	1569
Qy	1201	GlnMetArgArgGluArgLysAsnTyrIleuHisLeuGlnThrThrLysArgIleGln	1220
Db	1570	AGAGTAAAAACAGAAAGAAAACTTTTTCAGCAGTTCAGAAATCTGTACGAACCTATTTCAG	1629
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys	1240
Db	1630	GCTGCTTTTAGAGCATGAAA-----GTTAGACAAAAATTTGAAAAATGTATCAGAGAA	1683
Qy	1241	LysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
Db	1684	AAGATGGCAGCAATGTT-----AACCAATCTGCACCTCTGCTGTTCAGAAAGTAAA	1734
Qy	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisIleGln	1280
Db	1735	ACTCAGTATGAAGCTGTTCAAAGTGAAGGTGTATGATTCAAGAGTGTGTATAAGCTTCT	1794
Qy	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu	1300
Db	1795	GGCTTGCTGTTTCACAGAAAGCAGAGTATCATCTTCAAAGTAGGCTGTCAGTAACAAT	1854
Qy	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
Db	1855	CAAAAAGCTTTTGTAGATGCTGCACAGAAAACTGGAAACACAGAAATGTCTGCCCTA	1914
Qy	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
Db	1915	CGGATTCAGTTCTTCTTTCAGATGCTGTGTATCGGAGAAGATTTGTTTCAGCAAGAAAA	1974
Qy	1323	AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLys	1342
Db	1975	GCTGCTATCATTTCAGCAATTTTATTAGGAG-----TGGCAACACCAAGAAA	2022
Qy	1343	SerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
Db	2023	CAGTTTTTACTATATAGAAAAAGCAGCTGGTTTTTACAAAATCACTACAGCAATTTCTG	2082
Qy	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
Db	2083	TCTGCAAAACATCAAGCAAGCAAGTCTTATTACAGATCAGAAAGCAGTGTATTATTCAA	2142
Qy	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
Db	2143	GCTGAAGTAAGAGATTTTATACAGAAACGGAG-----TTTCAGGAATTTAAA	2190
Qy	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
Db	2191	AATAGCACCAATAAAATTCAGGCTATGTGGAGGAGATATAGACCAAGAAATTTATGCT	2250
Qy	1422	Lys-----AlaGlyTyrArg-----	1426
Db	2251	AAAGTGAAGGCTGCTCGCAAGATTCAAGCTGTGTATAGATTGTGGAGAGCACACAAAGAA	2310

QY 1426 ----- 1426  
 Db 2311 TATCTAGCTATATTAAAGCTGTATAAATATTCAAGTGTCTTCTATACAAACTAGAG 2370  
 QY 1427 -----ArglleArgLeuSerSerleuSerlleGlnArgLysTrpArgAla 1441  
 Db 2371 AGAACACGGTTTTTGAATGTGAGAGCATCAGCAATTATCATTCAGAGAAAATGAGAGCT 2430  
 QY 1442 ThrValGlnAlaArgGlnArgGluillePheLeuSerThrIleArgLysValArgLeu 1461  
 Db 2431 ATACTTCCTCAAGATAGCTCATGAACACTTCTTAATGATAAAAGA----- 2478  
 QY 1462 MetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMet 1481  
 Db 2478 ----- 2478  
 QY 1482 LysArgArgAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLys 1501  
 Db 2479 ---CATCGAGCTGCTGTGTGATCCAGACCATTTATAGAGGA-----TATAAA 2523  
 QY 1502 AlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArg 1521  
 Db 2524 GGAAGCAGCTCTTTCTCGCAGAAATCTGCTGTGTATCATCAAAAATATATACGA 2583  
 QY 1522 AlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaVal 1541  
 Db 2584 GCCAGGAGGCTGGAAGCATGAAGATAAAATATATATTGAAATTTAAAAAATCTACAGTT 2643  
 QY 1542 HisLeuGlnGlnLysPheArgGly-----LysArgLeuMetIleGluGlnArgAsn 1558  
 Db 2644 ATCCTTACAGCACTGGTGGCGTGTGGCTAGTAGTACGAAAGATTTTGTAGAACAGAGAGCC 2703  
 QY 1559 CysPheGlnLeuLeuArg----- 1564  
 Db 2704 AAAATTCGACTTCTTCACTTCAGCTGCAGCTGCATATATTACCTGAATGCTGTAGAAAT 2763  
 QY 1564 ----- 1564  
 Db 2764 CAAAGAGCCTATAAACTTTTACCTGGCTGTGAAGATGCTAAACAGCAGGTTAATTCAGTC 2823  
 QY 1565 ---CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
 Db 2824 ATCTGTATTTCAGATGGTGTTCGAGCAAGA-----TTACAAGAAAAGAGATTT 2871  
 QY 1584 ---GlnAlaLeuMetThrProGluMetMetAsp-----LeuIleArg 1596  
 Db 2872 ATTCAGAAATATCATAGATCAAAAGATTTGACATGACATGACATGCAAGATGCTGAGCCAG 2931  
 QY 1597 GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArg 1616  
 Db 2932 CGAAATAGGGCTGCATCAGTAAATACAGAAAGCAGTGCCTATTTCTCTCCGTAAAGAG 2991  
 QY 1617 GlnLysHis-----GlnGlyLeuLeuAspIleArg----- 1626  
 Db 2992 CAGGAAAAATTCATAGTGAATCATATAAAATTCAGGCAATTATGAGAGAGCTATTCTTGG 3051  
 QY 1627 -----LysArgIleAlaGlnLeuGlnGlnAlaLysAlaValAsn 1640  
 Db 3052 AGGAGAAAATCATTTGTACAAAATTAAGCTATACGACTAAGTCTTCAAGTTGTTAAT 3111  
 QY 1641 SerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSer 1660  
 Db 3112 AGG-----GAGATTTCAGAGAAAAACAACACTCTACAAAAGAACTGCACCTGCACCT 3162  
 QY 1661 AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu 1680  
 Db 3163 CATTACCTTTTGACATATAGCAACCTTTCTGCCATTTCTGAGGCCCTTAAACACCTAGAG 3222  
 QY 1681 MetTyrCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg 1700  
 Db 3223 GTAGTTTACAGA-----TTGCTCCACTTTGTTGTGAGAACATGCCAGAGTGGAGCA 3276  
 QY 1701 SerGluValAspLysGlnLeuIleGluArgCysSerArg----- 1713

Db 3277 ATTTCTAAATATTGTTGTTTGTATCCGAAGTTGTAATCGCAGTATTCCTTGTATGAAGTC 3336  
 QY 1714 -----llelleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsn 1727  
 Db 3337 ATCAGATATGCTGCAAGTCTTGTCTTAATGTATCTAAAGTATGAGAAAACCTACTTCAGCA 3396  
 QY 1728 ThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
 Db 3397 GTTTATGATGTAGAAAATGTATAGATATACTATTGGAGCTTTTGCAGATATACCGAGAA 3456  
 QY 1743 -----ArgTrpCysAspLysAspSerGluillePheAsnThrLeuCysThrLeu 1758  
 Db 3457 AAGCCTGGTAATAAGATTGCAGACAAAGCGGAGCATTTTACAAAAAATGTTGTTGTTG 3516  
 QY 1759 IleTrpValPheAlaHisCysProLysLysArgLysIlelleHisAspTyrMetThrAsn 1778  
 Db 3517 TTGGCTATTTTACTG-----AAGACAACAAAATAGAGCCTCTGATGTACGAAGTAGG 3567  
 QY 1779 ProGluAlaIleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMet 1798  
 Db 3568 TCCAAAGTTGTTGACCGTATTTACAGTCTCTACAACTTACAGCTCATAAACATAAATG 3627  
 QY 1799 LysGlnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys--- 1817  
 Db 3628 ---AATACTCAAAAGATA-----CTTTACAAGCAAAAGAGAAT 3663  
 QY 1818 -----lleAsnPheThrProCysSer-----LeuProSer 1827  
 Db 3664 TCTTCTATAGCAATCTCTTTTATCCAGAAACACCTGTAGGACCAGAAATAGTTTCAAGA 3723  
 QY 1828 LeuGluProAspPheGlyllelleArgTyrSerProTyrThrPheIleSerSerValTyr 1847  
 Db 3724 CTTAAGCCAGATGGGTTTGTGACAGAGATACATGAGAAATACAAAATCCCTGCAA 3783  
 QY 1848 AlaPheAspThrIleLeuCysLysLeuGlnIle 1858  
 Db 3784 GCTATTCAAATGGTGTGATGATGATGCGTTGGCATT 3816  
 RESULT 6  
 ABQ99318  
 ID ABQ99318 standard; cDNA; 4027 BP.  
 XX  
 AC ABQ99318;  
 DX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human coding sequence SEQ ID 51.  
 XX  
 KW Human; expressed sequence tag; EST; chromosome 17;  
 KW haematopoietic disorder; central nervous system disease; viral infection;  
 KW peripheral nervous system disease; non-healing wound; infectious disease;  
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;  
 KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;  
 KW antiallergic; antiinflammatory; immunosuppressive; fungicide;  
 KW cytostatic; haemostatic; virucide; antibacterial; neuroprotective;  
 KW immunostimulant; cerebroprotective; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200259260-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 16-NOV-2001; 2001WO-US042950.  
 XX  
 PR 17-NOV-2000; 2000US-00714936.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;  
 PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 XX

DR WPI: 2002-590824/63.  
 DR N-PSDB; ABP64732.  
 XX  
 PT New isolated polynucleotide, useful in research, diagnostic or  
 PT therapeutic methods, e.g. preventing or treating disorders involving  
 PT aberrant protein expression or biological activity.  
 XX  
 XX  
 XX Claim 1; SEQ ID NO 51; 394pp; English.  
 PS  
 CC The present invention relates to novel human coding sequences (ABQ99268-  
 CC ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in  
 CC therapeutic, diagnostic and research methods. The polynucleotides may be  
 CC used in the field of molecular biology as hybridisation probes, primers  
 CC for PCR, for chromosome and gene mapping, for the recombinant production  
 CC of protein, or in generation of anti-sense DNA or RNA. The  
 CC polynucleotides are useful in diagnostics as expressed sequence tags  
 CC (ESTs) for identifying expressed genes or for physical mapping of the  
 CC human genome. The proteins may be used as molecular weight markers, or as  
 CC nutritional sources or supplements. The proteins may be used to maintain  
 CC and expand cell population in a totipotent or pluripotent state  
 CC useful for re-engineering damaged or diseased tissues, transplantation,  
 CC manufacture of bio-pharmaceuticals or the development of bio-sensors. The  
 CC polynucleotides and proteins are useful for preventing, treating or  
 CC ameliorating disorders involving aberrant protein expression or  
 CC biological activity, e.g. haematopoietic disorders, central/peripheral  
 CC nervous system diseases, mechanical and traumatic disorders, non-healing  
 CC wounds, immune deficiencies and disorders, infectious diseases caused by  
 CC viral, bacterial or fungal infection, autoimmune disorders, allergic  
 CC reactions and conditions, coagulation disorders, or cancer. The  
 CC polynucleotide sequences of the invention were assembled from ESTs  
 CC isolated mainly by sequencing by hybridisation, and in some cases,  
 CC sequences obtained from one or more public databases. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIFO at  
 CC ftp.wifo.int/pub/published\_pmt\_sequences  
 XX  
 SQ Sequence 4027 BP; 1492 A; 672 C; 770 G; 1093 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6,9e-41 Length: 4027  
 Score: 700.50 Matches: 275  
 Percent Similarity: 40.70% Conservative: 226  
 Best Local Similarity: 22.34% Mismatches: 412  
 Query Match: 7.36% Indels: 319  
 Gaps: 40

US-09-914-698-1 (1-1861) x ABQ99318 (1-4027)

QY 868 TrpGlnLeuIleTyrIysPheArgSerProLysPheHisAlaAlaThrValLeuGln 887  
 Db 359 TGGATGATTAAGAAAGGATGCGA-----GAGATGCAAGGGCTGCTACTTTCACCG 412  
 QY 888 LysTrpTrpArgArgHisTrpLeuHis-----Val 897  
 Db 413 TCTACTTTCAGATGATGACAGATTACATATACAGATATCAGGCTTTGAAAACAGGCGCTCCGTT 472  
 QY 898 ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet 909  
 Db 473 GTGATCCACAGCAATACCAAGCAATAGAGTGCAGAACTGCAGAGCGCATTTATCTC 532  
 QY 910 ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys 929  
 Db 533 AGCAAAAGACACTCTGCTGTGATCCTTCAGGCTGCATTTCAGGGGTATGAAACTAGAGA 592  
 QY 930 TyrValLysLeuPheLysTrpGluArgThrGlnAlaAlaIleLeuLeuGlnLysPheThr 949  
 Db 593 CATTTCAAGAGTATGCATTCC-----TCTGCAACCTTATTTCAGAGTAGGTTT 640  
 QY 950 ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln 969  
 Db 641 AGATCATTAATGTTGAGGAGAGAGATTTCATTTCCCTCAAAAAGCTACTATTTTGTTCAG 700  
 QY 970 ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg 989

Db 701 AGGAATATCGAGCCACCATTGTGCCAAACATAAATTTGTACCAATTTCTGCACTTAAGA 760  
 QY 990 GluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeu--- 1008  
 Db 761 AAGGAGCCATTACAAATCAGTCATCTTTACAGAACACTGATGTTAAAGAAAGAGTTACAA 820  
 QY 1009 -----LeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln--- 1023  
 Db 821 GAATGCAAGGGCTGCAGT-TCTCATTCAGGCTACTTTCAGGATGCACAGAAAAA 879  
 QY 1024 -----GlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThr 1036  
 Db 880 TATATTACATTTTCAGACTTGGAAACATGCTTCAATTTCAATTTACAGCAACATTATCAACA 939  
 QY 1037 TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe 1056  
 Db 940 TATAGAGCTGCAAAATTCGAAAGAGAAATATATATACAGCAATGCGATTCCTCTGTGGTT 999  
 QY 1057 ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGlu 1073  
 Db 1000 ATTCAGGCTGCATATATAAGGATGAAGCAGACAACTTTAAGGAGAAACACAAAGCT 1059  
 QY 1074 -----PheGlnGlnLeu 1077  
 Db 1060 TCTATCGTAATACAAAGCACCTTACAGAAATGTATAGGCAGTATTTGTTCTACCAAAAGCTT 1119  
 QY 1078 LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys 1097  
 Db 1120 CAGTGGCTGCAAAATCATACAGAAATATATAGCAGCAATATAAGGAGAAACACAGAAA--- 1176  
 QY 1098 AsnAlaAspTyrLeuAlaLeuArgSerValLeuLys-----ValGlnAla 1113  
 Db 1177 -----GTATTTCAACATGAACTTAAGAAAGAGACTTGTGTTGAGCA 1221  
 QY 1114 ---TyrArgLysAlaThrIleGlnMetArgIleAsp----- 1124  
 Db 1222 GGTTCAGGACATGCAACATAAAAAACAGATTTCAGGAACAGCACCGGCTGCCATTATT 1281  
 QY 1125 -----ArgAsnHisTyrTyrSerLeuArgLysAsn 1134  
 Db 1282 ATTCAGAAAGCAATTGTAAGACCTTTAAAAATAAGGAGCAATATCTCCACCTTAGAGCAACA 1341  
 QY 1135 ValIleCysLeuGlnArgLeuArgAlaIleMetLysMetArgGlnGln----- 1151  
 Db 1342 GTAGTTTCTATTCAAGAGATACAGAAACTAACTGCGAGTCGCTACCCAGCAGTTATT 1401  
 QY 1152 -----ArgGluAsnTyrLeuArgLeuArgAsn 1160  
 Db 1402 TGTATACAGTCTTATTACAGAGGCTTTAAAGTACGAAAGGATATTCAAAATATGACCGG 1461  
 QY 1161 AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn 1180  
 Db 1462 GCTGCCACATAATTCAGTCATTCATTCGAAATGCACAGGGCCAAAGTTGAT----- 1512  
 QY 1181 AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu 1200  
 Db 1513 ---TATGAACAAAGAAACTGCATTTGGTGTATACAGAAATATATAGTTGTATGTT 1569  
 QY 1201 GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln 1220  
 Db 1570 AGAGTAAAAAATAAAAAAATACTTTTAGCAGTTTCAGAAATCTGTACCAACTATTTCAG 1629  
 QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLys 1240  
 Db 1630 GCTGCTTTTAGAGGATGAGAA-----GTTAGCAAAAAATTCAAAAATGATCAGAGGAA 1683  
 QY 1241 LysValThrLeuValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg 1260  
 Db 1684 AAGATGGCAGCCATTGTT-----AACCAATCTGCATCTGCTGTTACAGAACCAAA 1734  
 QY 1261 GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln 1280



Qy 1848 AlaPheAspThrIleLeuCysLysLeuGlnIle 1858  
 Db 3784 CTTATTCAAAGGCTGATGATACGCTTGGCAAT 3816

## RESULT 7

ABV20287/c  
 ID ABV20287 standard; cDNA; 1115 BP.

XX AC ABV20287;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 20278.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PS Claim 1; Page 3317-3318; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservatives: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV20287 (1-1115)

Qy 638 lleLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValVal 657  
 Db 1092 GTTCCTGAATTGGCTGTGTTCCTACAAATCCTTTGGCTTCGAATTGGCTACAGACAACT 1033

Qy 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db 1032 TATGGAGAACTCATATCTTTGGAGATAAACAGTGATGTCACAGGCTTGGCTATGTTATT 973  
 Qy 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTTGGAAATCCTGATATAGCAGCTGAGTATAGACACCCCACTGTTCT 913  
 Qy 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 Db 912 CACCTGTATAGAGATGGTCATGAAGAACTTTGTCCAACTTTACATTTGAAAAGTTATTG 853  
 Qy 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 Db 852 TTGTTGGCTCTGTTTCTTGAATATGCTAAATTTCCAGACTCATGTGATCATGCTTGT 793  
 Qy 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCCTCTCTAAAGATGCGCAATTCAGGCTAGTAAAGAAATCCTTTTGGCTTTTTCACGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgGluGlyTyrValLeu 774  
 Db 732 GATTTCCCTAAGTGGTGAAGGTGACCTTTCCCGTCACTTGGCTTAATGGGATTACCTGTT 673  
 Qy 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTTTCAGACACACCATTTGATGAATTTGATTTTGGCTTACAAATCTTGGCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGGAGTGGCTTGTGCGAACCACTGGAATCTCTCACACAGAACTGGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAAACTCAGGATTCGGCAATAAGTCTCTTCARAGATGCACAATGTTGACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGAGAAATTGAATTAAGTGATGATGAGCAATAACAATT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuTrpGlnLeu 870  
 Db 432 CTATCTAAGGATATTGTGATAGGCACAGAGAAAAAACTCTCAGATTGCTTTGGAAAAATA 373  
 Qy 871 lleTyrLysPheArg 875  
 Db 372 GCGTTTGTCTTTTCAG 358  
 RESULT 8  
 ID ABV20247/c  
 XX ABV20247 standard; cDNA; 1115 BP.  
 AC ABV20247;  
 XX 13-SEP-2002 (first entry)  
 DE Human prostate expression marker cDNA 20238.  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 OS Homo sapiens.  
 PN WO200160860-A2.  
 XX 23-AUG-2001.  
 XX 20-FEB-2001; 2001WO-US005171.  
 XX 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.







Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
pharmacogenomic marker; gene; ss.

KW  
XX  
XX  
OS  
XX  
PN  
XX  
XX  
PD  
XX  
PF  
XX  
PR  
PR  
PR  
PR  
PR  
PR  
PR  
XX  
XX  
PA  
PI  
XX  
DR  
XX  
XX  
PT  
PT  
PT  
XX  
FS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

Alignment Scores:  
Pred. No.: 4,06e-30 Length: 1115  
Score: 542.50 Matches: 111  
Percent Similarity: 64.90% Conservative: 48  
Best Local Similarity: 45.31% Mismatches: 79  
Query Match: 5.70% Indels: 7  
DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV26078 (1-1115)

QY 638 lleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuValVal 657  
Db 1092 GTCTGAATGGCTGTGCTCACTACATCTTTGGCTTGGATTCGATGCTAGAGCAACT 1033  
QY 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 677  
Db 1032 TATGAGAACTCATATCTTTGGAGATAACAGTGTACAGGGTGGCTATGTTTATT 973  
QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
Db 972 CTGAATCGCTACTTTGGAACTCCTGATATAGCAGCTGATATAGACACCCCACTGTCTCT 913  
QY 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
Db 912 CACCTGTATAGAGATGGTCATGAGAGCTTTGTCCAGTTTACATTTGAAAAGTATTG 853  
QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
Db 852 TTGTTGGTCTGTCTTCTGATATGCTAAATTTCCAGACTCATTTGATCATGCTCTGT 793  
QY 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
Db 792 CTCCTTGTAAAGATCCGAAATTCAGGCTAGTAAAGAAATCTTTGGCTTTTTCACGA 733  
QY 755 LeuLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeu 774  
Db 732 GATTTCTTAAGTGGTGAAGTGACCTTTCCCGTCACCTTGGCTTATTTGGATTAAGTCT 673  
QY 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
Db 672 AACCATGTTTCAGACACCAATTTGATGAATTTGATTTTGGCGTTACAAATCTTCCGTAGAC 613  
QY 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
Db 612 TTGCAATGGAGTGCGCTTTGCGAACCATGGAACCTTCTCACAGAACTGGGACCTC 553  
QY 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
Db 552 TCAAGAAACTCAGGATTCGGCAATAAGTCGTCTTCAAAGATGACAAATGTTGACATT 493  
QY 835 AlaLeuGlyAlaLeuGlyGluAlaLeuPheGlnLeu-----GlyGlyAspIle 850  
Db 492 GTTCTCAAGTCTTAAATCACGAGGAATTTGAATTAAGTGTAGCATGGAATAACAATT 433  
QY 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
Db 432 CTATCTAAGATATTGTGATAGGCACAGAGAAAACCTCTCAGATGCTTTGGAAAAATA 373  
QY 871 IleTyrLysPheArg 875  
Db 372 GCGTTTGTCTTTCAG 358  
RESULT 10  
ID ABV20220/c  
XX ABV20220 standard; cDNA; 1115 BP.  
XX AC  
XX ABV20220;  
XX 13-SEP-2002 (first entry)  
XX Human prostate expression marker cDNA 20211.  
XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 3304; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,06e-30 Length: 1115  
Score: 542.50 Matches: 111  
Percent Similarity: 64.90% Conservative: 48  
Best Local Similarity: 45.31% Mismatches: 79  
Query Match: 5.70% Indels: 7  
DB: 5 Gaps: 2

US-09-914-698-1 (1-1861) x ABV20220 (1-1115)

QY 638 lleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuValVal 857  
Db 1092 GTCTGAATGGCTGTGCTCACTACATCTTTGGCTTGGATTCGATGCTAGAGCAACT 1033  
QY 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 677  
Db 1032 TATGAGAACTCATATCTTTGGAGATAACAGTGTACAGGGTGGCTATGTTTATT 973  
QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
Db 972 CTGAATCGCTACTTTGGAACTCCTGATATAGCAGCTGATATAGACACCCCACTGTCTCT 913  
QY 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
Db 912 CACCTGTATAGAGATGGTCATGAGAGCTTTGTCCAGTTTACATTTGAAAAGTATTG 853  
QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734

Db 852 TTGTTGGTCTGTTTCTTGTATGATGCTAAATTTCCAGACTCATTTGATCATGATCCTTGT 793  
 Qy LeuPheVallyLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCTTCTGTAAGATGCGAATTCAGGCTAGTAAAGAAATCTTTGGCTTTTTCAGGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGTGAAGTGACCTTCCCGTCACCTTGGCTTATTGGGATTACCTGTT 673  
 Qy 775 GluHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTTTCAGACACCATTTGATGAATTTGATTTGCCGTTACAAATCTTTGCCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGAGTGGCTTGTGGACCATGGAATTCACAGAACTGGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAACTCAGGATTCGGCAATAAGTCTCTTCAAAAGATGCACAAATGTTGACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGGAATTCGATTAAGTGTGATGAGCATGGAATACAAAT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTATCTAAGGATATTGGATAGCACAGAGAAATACTTCAGATTGCTTTGGAAATA 373  
 Qy 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGTCTTTTCAG 358  
 RESULT 11  
 ID ABV26049/c  
 XX ABV26049 standard; cDNA; 1115 BP.  
 AC ABV26049;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 26040.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W0200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 5269; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX  
 SQ Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4.06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2  
 US-09-914-698-1 (1-1861) x ABV26049 (1-1115)  
 Qy 638 IleLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuValVal 657  
 Db 1092 GTCTGTAATGGCTGTTGTTCTCAATCCTTTGGGCTTCGAATGGTCTAGACAACT 1033  
 Qy 658 PheGlyLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerPheIle 677  
 Db 1032 TATGGAGAACTCATATCTTTGGAGAGATAACAGTATGTCACAGGTTGGCTATCTTATT 973  
 Qy 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTGGAAATCTCTGATATGACGCTGATATGACACCCCACTGTTCTCT 913  
 Qy 695 TyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeu 714  
 Db 912 CACCTGTATAGATGTCATGAAGAAGCTTTGCCAAGTTTACATTTGAAATAAAGATTATG 853  
 Qy 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCys 734  
 Db 852 TTGTTGGTCTGTTTCTTGTATGATGCTAAATTTCCAGACTCATTTGATCATGATCCTTGT 793  
 Qy 735 LeuPheVallyLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSer 754  
 Db 792 CTCTTCTGTAAGATGCGAATTCAGGCTAGTAAAGAAATCTTTGGCTTTTTCACGA 733  
 Qy 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTTAAGTGTGAGTGGCTTGTGGACCTTTCCCGTCACCTTGGCTTATTGGGATTACCTGTT 673  
 Qy 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAsp 794  
 Db 672 AACCATGTTTCAGACACCATTTGATGAATTTGATTTGCCGTTTACAAATCTTTGCCGTAGAC 613  
 Qy 795 LeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGAGTGGCTTGTGGACCATGGAATTCCTCACAGAACTGCGGACCTC 553  
 Qy 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAACTCAGGATTCGGCAATAAGTCTCTTCAAAAGATGCACAAATGTTGACATT 493  
 Qy 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTCTTAAATCAGGGAATTCGATTAAGTGTGATGAGCATGGAATACAAAT 433  
 Qy 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTATCTAAGGATATTGGGATAGCACAGAGAAATACTTCAGATTGCTTTGGAAATA 373

QY 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGCTTTTCAG 358  
 RESULT 12  
 ID ABV26118/c  
 XX ABV26118 standard; cDNA; 1115 BP.  
 AC ABV26118;  
 XX  
 XX 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 26109.  
 XX  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 5283; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
 CC determining whether prostate cancer has metastasized in a patient; (h)  
 CC assessing the aggressiveness or indolence of prostate cancer in a patient  
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
 XX  
 SQ Sequence 1115 BP; 357 A; 226 C; 205 G; 327 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 4,06e-30 Length: 1115  
 Score: 542.50 Matches: 111  
 Percent Similarity: 64.90% Conservative: 48  
 Best Local Similarity: 45.31% Mismatches: 79  
 Query Match: 5.70% Indels: 7  
 DB: 5 Gaps: 2  
 US-09-914-698-1 (1-1861) x ABV26118 (1-1115)  
 QY 638 IleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGluValVal 657  
 Db 1092 GTCCCTGAATTGGCTGTGTTCCTCAACATCTTTTGGCTTCGAATTTGGTCTAGAGACAAC 1033

QY 658 PheGlyGluLysIleGlnMetClnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
 Db 1032 TATGGAGAACTCATATCTTTGGAGATAACAGTGAATGTCACAGGGTTGGCTATGTTATT 973  
 QY 678 LeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLysAla 694  
 Db 972 CTGAATCGCTACTTTTGGAAATCCTGATATAGCAGCTAGTATAGACACCCACTGTCCT 913  
 QY 695 TyrThrLeuThrGluGluThrAlaGluThrIleLysLysHisSerLeuGlnLysLeu 714  
 Db 912 CACCTGTATAGAGATGTCATGAAGAAGCTTGTCCAAAGTTTACATGAAAAAGTTATTG 853  
 QY 715 PheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisnProCys 734  
 Db 852 TTGTGGTCTGTTCCTTTCCTTATGCTAAATTTCCAGACTCATTGATCATCTCTGT 793  
 QY 735 LeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 754  
 Db 792 CTCCTCTGAAGATGCCGAATTCAGGCTAGTAAGAAATCCTTTTGGCTTTTTCACGA 733  
 QY 755 GluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeu 774  
 Db 732 GATTTCTAAGTGTGAAGGTGACCTTTCCCGTCACCTTGGCTTATGGGATTACCTGT 673  
 QY 775 GlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeuAlaValAsp 794  
 Db 672 AACCATGTCAGACACCAATTTGATGAATTTGCGGTTTACAAATCTTGCCTAGAC 613  
 QY 795 LeuArgAspGlyValArgLeuThrArgValGluValIleLeuLeuArgAspLeu 814  
 Db 612 TTGCAATGTGGAGTGGCTTGTGGCAACCATGGAACTTCTCACAGAACTGGGACCTC 553  
 QY 815 ThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeu 834  
 Db 552 TCAAGAAACTCAGGATTCGCGCAATAAGTCGTCTTCAAAAGATGCACAATCTGCACAT 493  
 QY 835 AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeu-----GlyGlyAspIle 850  
 Db 492 GTTCTTCAAGTTCTTAAATCAGGAATTAAGTAAAGTATGATGATGATGGAATCAAT 433  
 QY 851 AlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeu 870  
 Db 432 CTATCTAAGGATATGTGATAGGACAGAGAAAAAATCTTCAGATTGCTTTGGAATA 373  
 QY 871 IleTyrLysPheArg 875  
 Db 372 GCGTTTGCTTTTCAG 358  
 RESULT 13  
 ABV09513  
 ID ABV09513 standard; cDNA; 752 BP.  
 XX  
 AC ABV09513;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 9504.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US005171.  
 XX  
 PR 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.





Qy	1524	ArgSerMet	LysGln	AlaArgGln	GluPheValGln	LeuArgThr	IleAlaValHis	Leu	1543
Db	1247	TTACAGAA	AAAGAGATTTATTTC	CAAAATATCAT	CATGATC	----	----	1285	
Qy	1544	GlnGln	LysPheArg	GlyLysArg	LeuMetIle	GlnArgAsn	CysPheGln	Leu	1563
Db	1286	-----	AAAAAGATTGAGCATG	GAAGTCA	GAGATGCT	CAGATGCT	CTGAGCCAGCGA	1327	
Qy	1564	ArgCys	SerMetPro	GlyPheGln	AlaArgAla	ArgGlyPhe	MetAlaArgLys	ArgPhe	1583
Db	1328	AATAGG	CGTCATCAGTAATA	TACAGAA	CGCAGTCGCC	CAATTTCT	CTCCCTCGTAAAAAG	CAG 1387	
Qy	1584	GlnAla	LeuMetThrPro	GluMetMet	AspLeuIle	ArgGlnLys	ArgAlaLys	Val 1603	
Db	1388	GA AAAA	TTCACTAGT-----	GGAAATCA	TTAAAA-----	-----	-----	1414	
Qy	1604	IleGln	ArgTyrTrp	ArgGlyTyr	LeuIleArg	ArgArgGln	LysHisGln	GlyLeu 1623	
Db	1415	ATTCA	GGCATTATGGAGAG	GGCTATCTT	CGGAGGA	AAAAAATGATT	GTACAAAAAT	TAAA 1474	
Qy	1624	AspIle	ArgLysArg	IleAlaGln	LeuArgGln	GluAlaLys	AlaValAsn	SerVal 1643	
Db	1475	GCTATAC	GACTAAGTCTTCA	AGTTGTTTAAT	AGGAGATTC	CGAGA	GA AAAA	CAAACTCTAC 1534	
Qy	1644	CysLys	ValGlnGlu	AlaValArg	PheLeuArg	GlyArgPhe	IleAlaSer	AspAla 1663	
Db	1535	AAAAGA	ACTGCATCTGCCTTC	ATCCTTTG	CAATATAG	CACCTTTCTG	CCATCTCTT 1594		
Qy	1664	AlaVal	LeuSerGln	LeuAsp	ArgLeuSer	ArgThrValPro	HisLeuLeu	MetTrp 1683	
Db	1595	GAGGCCT	TAAACACACCT	TAGAGGT	AGTTTACT	TAGA-----	-----	1627	
Qy	1684	SerGlu	PheMetSerThr	PheCysTyr	GlyIleMet	AlaGlnAla	IleArgSer	GluVal 1703	
Db	1628	-----	TTGTCCTCCACTTT	GTGTGAGA	CAATGCGCC	AGAGTGAG	GCACAA	TTTCTPAAA 1678	
Qy	1704	AspLys	GlnLeuIle	GluArgCys	SerArg-----	-----	-----	1713	
Db	1679	ATA	TTTGTGTTGATCC	GAAGTTGAT	CGCAGTATCT	CTTGTATG	GAAGTCAT	CAGATAT 1738	
Qy	1714	-----	IleIleLeu	AsnLeuAla	ArgTyrAsn	SerThrThrVal	AsnThrPhe	Gln 1730	
Db	1739	GCTGT	CAAGCTCTTGCTT	TAATGTATG	ATTAAGAT	TATGAGAA	AACTACTCT	CAGCAGTTTATGAT 1798	
Qy	1731	GluGly	GlyLeuValThr	IleAla	GlnMetLeu	Leu-----	-----	1742	
Db	1799	GTAGAAA	ATTGTATAGATATA	CTACTATT	CGAGCTTTTG	CAGATATAC	CCAGAAA	AGCCTGGT 1858	
Qy	1743	---Arg	TrpCysAsp	LysAsp	SerGluIlePhe	AsnThrLeu	CysThrLeu	IleTrpVal 1761	
Db	1859	AATAA	AGTTGCAGACAA	AGGCGGAAG	CATTTTAT	CAAAAA	CTTGTTGTT	TGTTGGCTATT 1918	
Qy	1762	PheAla	HisCysPro	LysLysArg	LysIleIleHis	AspTyrMet	ThrAsnPro	GluAla 1781	
Db	1919	TTACTG	-----	AGACAC	AAATAGAG	CCCTCTGAT	TACGAAG	TAGTCCAAAGTT 1969	
Qy	1782	IleTyr	MetValArg	GluThrLys	LysLeuVal	AlaArgLys	GluLysMet	LysGlnAsn 1801	
Db	1970	GTTG	ACCGTATTTAC	GTCTCTAC	AACTTTAC	AGCTCAT	ATAACATA	AAATG-----AAT 2023	
Qy	1802	AlaArg	LysProPro	MetThrSer	GlyArgTyr	LysSerGln	Lys-----	-----	1817
Db	2024	ACTGAA	AGAAATA-----	CTTTACA	AGCAAA	AGAAATTTCT	CTATA 2065		
Qy	1818	---Ile	AsnPheThrPro	CysSer-----	-----	-----	-----	1830	
Db	2066	AGCA	TTCTTTATCC	CAGAAA	CACCTGT	AAGGC	CAAGATAG	TTTCAAGACTTAAGCCCA 2125	
Qy	1831	AspPhe	GlyIleIleArg	TyrSerPro	TyrThrPhe	IleSerSerVal	TyrAla	PheAsp 1850	
Db	2126	GATTGG	TTTTTGAG	AGAGAT	ATAACAT	GGAAGAA	ATCAAA	TCCCTCGAAGCTTATTCAA 2185	
Qy	1851	ThrIle	LeuCysLys	LeuGln	Ile 1858				

Db	2186	ATGGTGATGATACGCTTGGCATT	2209
RESULT 15			
ACC50994			
ID	ACC50994	standard; cDNA; 2414 BP.	
XX	ACC50994;		
AC	ACC50994;		
XX	12-JUN-2003	(first entry)	
DT	12-JUN-2003	(first entry)	
XX	Human bladder cancer associated cDNA s		
DE	Human; bladder cancer; cytostatic; gen		
XX	Human; bladder cancer; cytostatic; gen		
KW	Human; bladder cancer; cytostatic; gen		
XX	Homo sapiens.		
OS	WO2003003906-A2.		
PN	WO2003003906-A2.		
XX	16-JAN-2003.		
PD	16-JAN-2003.		
XX	03-JUL-2002; 2002WO-US021338.		
PF	03-JUL-2002; 2002WO-US021338.		
XX	03-JUL-2001; 2001US-0302814P.		
PR	03-AUG-2001; 2001US-0310099P.		
PR	08-NOV-2001; 2001US-0343705P.		
PR	13-NOV-2001; 2001US-0350666P.		
PR	12-APR-2002; 2002US-037246P.		
XX	(EOSB-) EOS BIOTECHNOLOGY INC.		
FA	(EOSB-) EOS BIOTECHNOLOGY INC.		
XX	Mack DH, Aziz N;		
PI	Mack DH, Aziz N;		
XX	WPI: 2003-201532/19.		
DR	P-PSDB; ABR48183.		
DR	P-PSDB; ABR48183.		
XX	Detecting a bladder cancer-associated		
PT	patient, comprises contacting a biolog		
PT	bladder cancer-associated polynucleoti		
XX	Claim 6; Page 261; 307pp; English.		
PS	Claim 6; Page 261; 307pp; English.		
XX	The present invention describes a meth		
CC	associated transcript in a cell from a		
CC	contacting a biological sample from th		
CC	that selectively hybridises to a sequ		
CC	table of sequences (see ACC50951 to		
CC	encode the human bladder cancer-associ		
CC	ABR48242). Bladder cancer-associated s		
CC	have cytostatic activities, and can be		
CC	in vaccine production. The method can		
CC	cancer-associated transcript in a cell		
CC	compounds that modulate bladder cancer		
CC	useful in diagnosing or treating blad		
CC	The nucleic acid molecules from the pr		
CC	various screening and diagnostic metho		
CC	and/or antisense/inhibition applicatio		
XX	Sequence 2414 BP; 859 A; 391 C; 476 G;		
SQ	Sequence 2414 BP; 859 A; 391 C; 476 G;		
Alignment Scores:			
Pred. No.:	5.97e-22	Leng	
Score:	437.00	Match	
Percent Similarity:	36.74%	Cons	
Best Local Similarity:	21.26%	Mism	
Query Match:	4.59%	Indel	
DBs:	7	Gaps	
US-09-914-698-1 (1-1861) x ACC50994 (1-2414			
QY	913 ArgAlaAlaThrValIleGlnAlaValph		
Db	2 AAATGTGTACGAACACTTACGCTGCTTT		

QY 933 LeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyr 952  
Db : : : : :  
62 AATGATATCAGAGGAAAG---ATGCGACCCATTGTTAACCAATTCGCACTCTGCTGTATC 118  
QY 953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr 971  
Db : : : : :  
119 AGAAGCAAACTCAG---TATGAGCTGTTCAAGTGNAGGTGTATGATTCAGAGTGG 175  
QY 972 TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla 991  
Db : : : : :  
176 TATAAAGCTCTGCTGCTTGTGTTTCACGAAAGCAGAGTATCATCTCTCAAAAGTAGGGCT 235  
QY 992 AlaIlePheLeuGlnArgIleTrpArgArgGlnPheAlaLysLysLeuLeuAlaAla 1011  
Db : : : : :  
236 GCAGTAACAATTCAAAAGCTTTT---TGTAGAATGGTCACAGAAACTG--- 283  
QY 1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaLaserTyrIleGln 1031  
Db : : : : :  
284 ---GAAACACAGAAATGCTGCTCCCTACGGATTCAG 316  
QY 1032 ---MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg 1049  
Db : : : : :  
317 TTCTTCTTCAGATGGCTGTAT---CGGAGAAATTTGTTTCAG 358  
QY 1050 GlnArgAspLeuIleMetPheValGlnArgMetArgSerLysTyrSerMetLeuGlu 1069  
Db : : : : :  
359 CAGAAAAGAGCTGCTCACTTACAGCATTTATTTAGAGCG---TGG---CAA 406  
QY 1070 GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnArgTyrArg 1089  
Db : : : : :  
407 ACCGAAACAGTTTTTACTATATAGAAAAGCAGCGTGGTTTACAAAATCACTACAGA 466  
QY 1090 AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu 1109  
Db : : : : :  
467 GCATTTCTGCTGCACAAACATCAAGACAGAGCTATTTTACAGATCAGAAAGCAGTGTATC 526  
QY 1110 LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyr 1129  
Db : : : : :  
527 ATTATTCAAGCTAGAAGTAAAGGATTTATACAGAAACGG--- 565  
QY 1130 SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg 1149  
Db : : : : :  
565 --- 565  
QY 1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln 1166  
Db : : : : :  
566 ---AAGTTTCAGGAAATTAATAATAGCACCATATAAAATTCAGGCTATGTGG 613  
QY 1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186  
Db : : : : :  
614 AGGAGATATAGACCAAGAAA---TATTTATGTAAGTG 649  
QY 1187 LysCysIleIleAsnValGln---ArgArgTyrArgAlaThrLeuGlnMetArg 1203  
Db : : : : :  
650 AAAGCTGCTGCAGAAATTCAGCTGTATAGATGTTGAGAGCA--- 694  
QY 1204 ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnLysPhe 1223  
Db : : : : :  
695 ---CACAAAAGATATCTAGCTGTATTAAAAGCTGTATAAATTTATTCAGGTTGCTTC 748  
QY 1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThr 1243  
Db : : : : :  
749 TATACCAAACTAGAG---AGAACACGGTTTTTTGAATGTGAGACATCAGCA 796  
QY 1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263  
Db : : : : :  
797 ATTATCAITTCAGAAAATGGAGCTATCTCTCGCAAGAGTAGCTCATGACACTTC 856  
QY 1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgPheHisAlaGlnLysSerMet 1283  
Db : : : : :  
857 TTAATGATAAAAGA--- 871

QY 1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis 1303  
Db : : : : :  
872 ---CATCGAGCTGCTTGTGTTGATCCAGCACAT 901  
QY 1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323  
Db : : : : :  
902 TATAGAGCA--- 910  
QY 1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343  
Db : : : : :  
911 ---TATAAGGAAGGAGGCTCTCTCTTCGGCAG--- 940  
QY 1344 TyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363  
Db : : : : :  
941 ---AAATCTGCTGCTTTGATCATACAAAAATATATACGAGCAGGAGGCT 988  
QY 1364 MetGlnLysGlnValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg 1383  
Db : : : : :  
989 GGAAGCATGAAGGATAAATAATATTGAATTTAAAAAATCTACAGTTTATCTACAGCA 1048  
QY 1384 ArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGln 1403  
Db : : : : :  
1049 CTGTGCTGCTGCTGCTAGTAGTACGAAAAGATTTTAGAA---CAGAGA 1093  
QY 1404 AlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeuCysLysAla 1423  
Db : : : : :  
1094 GCCAAATTCACCTT---CTTCACCTCAGTCAGCT 1126  
QY 1424 GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrVal 1443  
Db : : : : :  
1127 GCATATTATACCTGAAT--- 1144  
QY 1444 GlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
Db : : : : :  
1144 --- 1144  
QY 1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArg 1483  
Db : : : : :  
1144 --- 1144  
QY 1484 ArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db : : : : :  
1145 ---GCTGTTAGAAATTCAAAGAGCCTATAAATTTACCTGGCTGTGAAGAATGCTAAC 1198  
QY 1504 GlnAspTyrGlnLeuLeuGlnSerSerValIleLeuValGlnArgCysPheArgAlaAsn 1523  
Db : : : : :  
1199 AAGCAGGTT---AATTCAGTCACTCTGTTTACAGATGGTTTCGAGCAAGA 1246  
QY 1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db : : : : :  
1247 TTACAAGAAAGAGATTTATTCAGAAATATCATAGCATC--- 1285  
QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db : : : : :  
1286 ---AAAAAGATTGAGCATGAAGGTCAAGAATGTCTGAGCCAGCA 1327  
QY 1564 ArgCysSerSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db : : : : :  
1328 AATAGGCTGCATCAGTAATACAGAAAGCAGTGCCTCATTTTCTCCTCGTAAAAGCAG 1387  
QY 1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaLysVal 1603  
Db : : : : :  
1388 GAAAAATTCACCTAGT---GGAATCATTA--- 1414  
QY 1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArgGlnLysHisGlnGlyLeuLeu 1623  
Db : : : : :  
1415 ATTCAAGCATTTATGAGAGGCTATTCTTGGAGAAAGAAAATGATGTACAAAAATTA 1474  
QY 1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg 1643  
Db : : : : :  
1475 GCTATACACTAAGTCTTCAAGTGTGTTAATGGGAGATTTCGAGAGAAACAACTTAC 1534  
QY 1644 CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663

```
Db      :::      |||:::      :::|||      |||      |||
1535 AAAGAACTGCACCTGTCACCTTACCTTTTGACATATAAGCACCTTTCTGCCATTCTT 1594
Qy      |||
1664 AlaValLeuSerGlnLeuSerArgThrValProHisLeuLeuMetTipCys 1683
Db      |||      |||:::      |||
1595 GAGCCTTAAACACCTAGAGTAGTTACTAGA----- 1627
Qy      |||
1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703
Db      :|||
1628 -----TTGTCCTCCACTTTGTTGTGAGAACATGGCCAGAGTGGAGCAATTTCTAAA 1678
Qy      |||
1704 AsplysGlnLeuIleGluArgCysSerArg----- 1713
Db      |||
1679 ATATTGTTTGTATCCGAAGTGTGAATCGCAGTATTCCTTGTATGGAAGTCATCAGATAT 1738
Qy      -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730
Db      :|||
1739 GCTGTGCAAGCTTGTCTTAAGTATCTAAGTATGAGAAACTACTTCAGCAGTTTAIGAT 1798
Qy      |||
1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742
Db      :|||
1799 GTAGAAAATTGTATAGATATACTATTGGAGCTTTTGCAGATATACCGAGAAAAGCCTGGT 1858
Qy      ---ArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyrVal 1761
Db      :|||
1859 AATAAAGTTGCAGACAAAGCGGAGCAATTTTACAAAACTTGTGTGTTTGTGGCTATT 1918
Qy      |||
1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781
Db      :|||
1919 TTAGTG-----AAGACACAAATAGAGCCTCTGATGTACGAAGTAGTCCAAAGTT 1969
Qy      |||
1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801
Db      :|||
1970 GTTGACCGTATTACAGTCTCTCAAACTTACAGCTCATAAACATAAATG-----AAT 2023
Qy      |||
1802 AlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817
Db      :|||
2024 ACTGAAGAATA-----CTTTACAAGCAAAAGAAGAAATTTCTTATA 2065
Qy      ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
Db      |||
2066 AGCATTCCTTTTATCCAGAAACACCTGTGAAGCAGACAGATAGTTTCAAGACTTAAGCCA 2125
Qy      |||
1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
Db      |||
2126 GATTGGGTTTGTGAGAGAGATAACATGGAAGAAATCACAATCCCTGCAAGCTATTCAA 2185
Qy      |||
1851 ThrIleLeuCysLysLysLeuGlnIle 1858
Db      :|||
2186 ATGGTGATGGATACGCTTGGCAATT 2209
```

Search completed: September 15, 2004, 02:07:39  
Job time : 1632 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 15, 2004, 00:52:24 ; Search time 9578 Seconds  
(without alignments)  
5802.205 Million cell updates/sec

Title: US-09-914-698-1  
Perfect score: 9514  
Sequence: 1 MELVWSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool\_h/US09914698/runat\_10092004\_150005\_28235/app\_query.fasta\_1.2055  
-DE=EST -QFMT=tastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09914698 @CGN 1.1.7670 -runat\_10092004\_150005\_28235 -NCPU=6 -ICPU=3  
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estini.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estrio.\*  
8: em\_hic.\*  
9: gb\_estcl.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estom.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_pro.\*  
26: em\_gss\_rtd.\*  
27: em\_gss\_phg.\*  
28: em\_gss\_vrl.\*  
29: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1391	14.6	1101	29	CNS00G5G	AL071458 Drosophil
2	1115	11.7	1022	29	CNS00LY6	AL078898 Drosophil
3	1002	10.5	596	9	A1519740	A1519740 LD39479.5
4	983.5	10.3	911	29	CNS015NX	AL105687 Drosophil
5	851	8.9	718	12	B1642052	B1642052 SD25436.5
6	723.5	7.6	3130	11	BC026699	BC026699 Mus muscu
7	716.5	7.5	3983	11	AK052326	AK052326 Mus muscu
8	647	6.8	589	12	BG636223	BG636223 SD13838.5
9	475	5.0	2675	11	AK048277	AK048277 Mus muscu
10	429	4.5	485	28	BH374743	BH374743 AG-ND-143
11	388	4.1	1008	29	CNS04QCR	AL302436 Tetraodon
12	359.5	3.8	526	9	AL720781	AL720781 AL720781
13	359.5	3.8	526	9	AL720782	AL720782 AL720782
14	334	3.5	619	28	CC129703	CC129703 ND1.19P17
15	329	3.5	852	29	CNS046XE	AL277259 Tetraodon
16	316	3.3	771	9	AU127218	AU127218 AU127218
17	309	3.2	983	13	BUI93761	BUI93761 AGENCOURT
18	304.5	3.2	755	13	BU222936	BU222936 603948053
19	291.5	3.1	789	13	EX077268	EX077268 EX077268
20	287	3.0	919	13	BQ952955	BQ952955 AGENCOURT
21	285	3.0	582	13	BU206367	BU206367 604155180
22	282.5	3.0	420	14	CD036256	CD036256 mgmc012xd
23	280.5	2.9	689	12	BGI77316	BGI77316 602313918
24	278.5	2.9	937	13	EX702765	EX702765 EX702765
25	277	2.9	592	9	AA539252	AA539252 LD18929.5
26	273	2.9	792	13	BU055628	BU055628 UI-M-P00-
27	270	2.8	839	13	BU420251	BU420251 603959350
28	269.5	2.8	696	13	EX087585	EX087585 EX087585
29	269	2.8	771	9	AL718233	AL718233
30	269	2.8	772	13	BM079393	BM079393 BM079393
31	268.5	2.8	773	14	CA381839	CA381839 661485 NC
32	267	2.8	704	12	BGI75972	BGI75972 602337893
33	267	2.8	725	14	CD350820	CD350820 UI-M-G10-
34	267	2.8	1131	13	EX383325	EX383325 EX383325
35	265	2.8	769	13	EX077267	EX077267 EX077267
36	263	2.8	856	13	BU911386	BU911386 AGENCOURT
37	262.5	2.8	664	28	AZ384962	AZ384962 IM0143011
38	257.5	2.7	819	12	BGI14414	BGI14414 602285705
39	255.5	2.7	4137	29	AY420057	AY420057 Homo sapi
40	253.5	2.7	955	29	CNS046XF	AL277260 Tetraodon
41	253.5	2.7	6294	29	AY410631	AY410631 Homo sapi
42	252	2.6	566	14	CF178179	CF178179 807218 MA
43	251.5	2.6	711	9	AL718189	AL718189 AL718189
44	249.5	2.6	6237	29	AY410633	AY410633 Mus muscu
45	248.5	2.6	829	14	CF289447	CF289447 AGENCOURT

# ALIGNMENTS

RESULT 1  
CNS00G5G  
LOCUS  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR32J06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL071458  
VERSION AL071458.1 GI:4951500  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1101)  
 Genoscope.  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osogawa and  
 Aaron Mammoss in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCL-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 p1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# REFERENCE

## AUTHORS

### TITLE

### JOURNAL

# COMMENT

# FEATURES

## source

1. 1101  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR32J06"  
 /clone\_lib="RPCL-98"  
 /note="end : TET3"

# ORIGIN

Alignment Scores:  
 Pred. No.: 1.94e-108 Length: 1101  
 Score: 1391.00 Matches: 280  
 Percent Similarity: 92.46% Conservative: 2  
 Best Local Similarity: 91.80% Mismatches: 3  
 Query Match: 14.62% Indels: 21  
 DB: 29 Gaps: 1

US-09-914-698-1 (1-1861) x CNS00GSG (1-1101)

QY 1487 ValValileGlnArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyr 1506  
 Db 4 GTAGTATCCAGSCCGTTTCGTCTCGTGTGCTATGTAAGGCGACAGGATTAC 63  
 QY 1507 GlnLeuileGlnSerValileLeuValGlnArgLysPheArgAlaAsnArgSerMet 1526  
 Db 64 CAATTAATCCAAATCCTCTGTGATCCTCGGCGAGCGCAAAATTCGGTGTCTAATCGCAGCATG 123  
 QY 1527 LysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLys 1546  
 Db 124 AAGCAGCAGCGCCAGGAATTTGTCACGTCGCTACTATTGCGATTCGATTTGCCAACAAAG 183  
 QY 1547 PheArgLysArgLeuMetIleGlnArgAsnCysPheGlnLeuLeuArgCysSer 1566  
 Db 184 TTCGTGGCGAAGCGTCTAATGATGAGCAGCGTAATTTTCCAACTGCTCCGCTGTTC 243  
 QY 1567 MetProGlyPheGlnAlaArgAlaArgLysPheMetAlaArgLysArgPheGlnAlaLeu 1586  
 Db 244 ATGCGCGGCTTCCAGGCACTGCGCGCGCTTTATGCTCGCAAAACGATTCGAGCCCTG 303  
 QY 1587 MetProGluMetMetAspLeuileArgGlnLysArgAlaAlaLysValileGlnArg 1606  
 Db 304 ATGACACCCGAGATGATGGACCTATCCGCGAAGCGCGCCGCGCAAGGTTATACCGT 363  
 QY 1607 TyrTipArgGlyTyrLeuileArgArgGlnLysHisGlnGlyLeuLeuAspIleArg 1626  
 Db 364 TACTGGCGGGCTATCTAATCCGACGCTCAGAACGACCGAGCGTCTCTTGGATATCCGG 423  
 QY 1627 LysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLysVal 1646  
 Db 424 AAGCGTATTGCTCAGCTGCGCAAGGCAAGCGGTAACCTCTGTGCGCTGCAAAAGTC 483

QY 1647 GlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaValLeu 1666  
 Db 484 CAGAGAGCGGTGCTTTCCTTCGCGACGCTTTATCGCATCATGCGTTACA-GTGGTA 542  
 QY 1667 SerGlnLeu----- 1669  
 Db 543 AGTCGATTGGTAAGTGATCCAGCTAGCTTAGTCTTATTATACTAATCCTCTTAATA 602  
 QY 1670 -----AspArgLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPhe 1686  
 Db 603 ATCCCGCGCATGCTCTTTTCGCGACTGTGCCACACTGCTCATGTGGTGTTCGGAGTTC 662  
 QY 1687 MetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGln 1706  
 Db 663 ATGTCACAGTTTGTCTATGGCATCATGCTCAGCCATTCGATCAGAGGTGGATAGCAG 722  
 QY 1707 LeuileGluArgCysSerArgIleLeuLeuAsnLeuAlaArgTyrAsnSerThrVal 1726  
 Db 723 CTTATAGAGCGCTGCAGCGCGATCATCTAAATTTGCGCGCTACCAATAGCACCGGTG 782  
 QY 1727 AsnThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeuArgTyrCysAsp 1746  
 Db 783 AACACGTTCCAGAGGCGGCTTTGGTCACCATTTGCCAGATGTTATTCGCTGGTGGAC 842  
 QY 1747 LysAspSerGluilePheAsnThrLeuCysThrLeuileTrpValPheAlaHisCysPro 1766  
 Db 843 AAGACAGTGTGATATTAACACTTTGTGACCCCTCATTTGGTATTCGCCCATCTGCT 902  
 QY 1767 LysLysArgLysIle 1771  
 Db 903 AAAAAACGAAAGGTA 917

# RESULT 2

# CNS00LY6

# LOCUS

# DEFINITION

# ACCESSION

# VERSION

# KEYWORDS

# SOURCE

# ORGANISM

# REFERENCE

# AUTHORS

# TITLE

# JOURNAL

# COMMENT

CNS00LY6 1022 bp DNA linear GSS 14-JUN-1999  
 Drosophila melanogaster genome survey sequence TET3 end of BAC:  
 BACR48008 of RPCL-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.

AL078898  
 AL078898.1 GI:5102188

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1022)

Genoscope.

Direct Submission

Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osogawa and  
 Aaron Mammoss in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCL-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 p1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. 1022

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

# FEATURES

## source

/clone="BACK48008"  
/clone.lib="RPCL-98"  
/note="end : FET3"

## ORIGIN

## Alignment Scores:

Pred. No.: 9,08e-85 Length: 1022  
Score: 1115.00 Matches: 228  
Percent Similarity: 82.6% Conservative: 6  
Best Local Similarity: 80.5% Mismatches: 28  
Query Match: 11.72% Indels: 21  
DB: 29 Gaps: 1

US-09-914-698-1 (1-1861) x CNS00LY6 (1-1022)

QY 1512 SerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGln 1531  
DB 1 TCTGARATCCTGGAGCAGCGCAAAATWCCGTGTCTAATCGCAGCATGAAGCAGCGCCAG 60  
QY 1532 GluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArg 1551  
DB 61 GAATGTGTCCAGCTGGTACTATTTCAGTCCATTGTCACCAAAAAATTCGTGGCAAGGT 120  
QY 1552 LeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGln 1571  
DB 121 CTAATGATTGARCARGTAATGTWTCACACWGCTCGCTGTTCATGCCGGGCTTCCAG 180  
QY 1572 AlaArgAlaArgGlyPheMetAlaArgLysPheGlnAlaLeuMetThrProGluMet 1591  
DB 181 GCACGTGCCGGCGGCTTTATGCTCGCAAAATAATCCAGGCCCTGAWGAMACCCGARAWA 240  
QY 1592 MetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyr 1611  
DB 241 AWGACCTCAWCCGCGCARAAGCGCGCCGCAAGGTAAACARCGTTACAGCGGGGCTAT 300  
QY 1612 LeuLeuArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLys-ArgIleAlaGln 1631  
DB 301 CTAACCCGAMGAMRTCAAAAGCACCAAGTCTCTTTGGATATCCGGAAGAGTATTACTCA 360  
QY 1631 nLeuArgGlnGluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValAr 1651  
DB 361 ACTGGGCAAAAGCAAAAGCGGTAAACWCATATGCGTGCACAAAGTCCAGGAGCGGTGCG 420  
QY 1651 qPheLeuArgGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeu- 1669  
DB 421 TTTCTTCGGGACGCTTTATCGCAATCAGATCGGTASAGTTGCTTAAGTCGATTGGGTAA 480  
QY 1670 -----AspAr 1671  
DB 481 GTGATCCAGCTAGTTAGTCTTATTATACTAATCTCTTAATAATCCCGCAGATCG 540  
QY 1671 qLeuSerArgThrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCy 1691  
DB 541 TCTTTCCGGGCACTGCGCACACCTCTCATGTGGTGTTCGGAGTTTCATGTCACGCTTTTK 600  
QY 1691 sTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCy 1711  
DB 601 CTAATGCAATCATGGCTAGGCAATTCGATCAGAGTGGATAGACGCTTATAGACGCTG 660  
QY 1711 sSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGlnGln 1731  
DB 661 CAGCGGATCATCTTAATKTBGCGCGCTACAAATAGCACCCAGCGGTGAACAGCTTCCASGA 720  
QY 1731 uGlyGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAspLysAspSerGluI 1751  
DB 721 GGGCGGTTTTGTCAACATTGGCCAGATGTATTGCGCTGTGTCGCAAAAGACAGTGAAT 780  
QY 1751 ePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysProLysLysArgLysI 1771  
DB 781 ATTCAASACTTTGTGACCTCATTTGGKATTCGCCCACTGTCTCTAAAAAGCGAAGKA 840  
QY 1771 eIleHis 1773  
DB : : : : :

Db 841 GCTAAAT 847

## RESULT 3

AI519740

## LOCUS

## DEFINITION

AI519740 596 bp mRNA linear EST 02-DEC-2003  
LD39479.5prime LD Drosophila melanogaster embryo pOT2 Drosophila  
melanogaster cDNA clone LD39479 5 similar to asp: FBan0006875  
GO: microtubule binding (GO:0008017); microtubule binding  
(GO:0008017); cytoskeletal structural protein (GO:0005200);  
microtubule associated protein (GO:0005875); microtubule located on:  
3R 96A23-96A23; 08/13/2002, mRNA sequence.

## ACCESSION

AI519740

## VERSION

AI519740.1 GI:4425594

## KEYWORDS

## SOURCE

## ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 596)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Other ESTs: LD39479.3prime

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd., Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

hit genomic AE003749: arm:3R [20443564,20672042]

estimated-cyto:96A20-96B10: 04/10/2001

Plate: LD.394 row: G column: 7

High quality sequence stop: 453

POLYA=No.

## FEATURES

Location/Qualifiers

1..596

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clone="LD39479"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="xhl Blue"

/clone\_lib="LD Drosophila melanogaster embryo pOT2"

/note="Organ: embryo; Vector: pOT2; Site 1: EcoRI; Site 2:

XhoI; Sized fractionated cDNAs were directly ligated into

pOT2."

## ORIGIN

## Alignment Scores:

Pred. No.: 1.85e-75 Length: 596

Score: 1002.00 Matches: 197

Percent Similarity: 99.49% Conservative: 0

Best Local Similarity: 99.49% Mismatches: 1

Query Match: 10.53% Indels: 0

DB: 9 Gaps: 0

US-09-914-698-1 (1-1861) x AI519740 (1-596)

QY 607 ProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSerAspArg 626

DB 1 CCATGCTCCAAAGTGGCGGTATATGTCAACAGCAAGCTCTGCGCATCCGTAGCGATCGT 60

QY 627 AsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLysCysPheAsn 646

DB 61 AATCTTACCTAGACGTAGTATTGCAACGCCACCATCTTGAGCTGCTGCTTCTCAAT 120

QY 647 ProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGluLysIleGlnMetGlnSer 666

DB 121 CCCCTTTGGCTGGCTTGGACTGGAGTGTCTTTGGCGAGAGATCCAGATGCAGTCT 180



```

REFERENCE
AUTHORS      1 (bases 1 to 718)
              Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
              Lewis,S. and Rubin,G.M.
TITLE        BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
              BDGP
              Lawrence Berkeley National Lab
              One Cyclotron Rd, Berkeley, CA 94720, USA
              Fax: 510 486 6798
              Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
              hit genomic AB003749: arm:3R [20443564,20672042]
              estimated-cyto:96A20-96B10: 05/19/2001
              Plate: SD.254 row: C column: 12
              High quality sequence scop: 675.
              Location/Qualifiers
                1..718
                  /organism="Drosophila melanogaster"
                  /mol_type="mRNA"
                  /db_xref="taxon:7227"
                  /clone="SD25436"
                  /lab_host="DHS-alpha"
                  /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                  culture pOT2"
                  /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
                  fractionated cDNAs were directly ligated into pOT2.
                  Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.:      2,446-62      Length:      718
Score:          851.00      Matches:    176
Percent Similarity: 85.8%      Conservative: 0
Best Local Similarity: 85.8%      Mismatches: 2
Query Match:    8.94%      Indels:    28
DB:            12          Gaps:        1

US-09-914-698-1 (1-1861) x B1642052 (1-718)

QY      1 MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuGlnLeuIle 20
DB      614 ATGGAGTAGTGTGGAGCNCCTCTCTTGGAGTGGCTGCAAGGAGCGCTGCAGCTAATA 555

QY      21 AspAsnArgAsnPheArgLysGluValMetIleLeuLysSerLysSerGlnPro 40
DB      554 GACACCGCACTTCGAAAGAGGGTGATGATCACTCAAGTCCAAAGACCAACGCGC 495

QY      41 Val----- 41
DB      494 GTCAAGGTCAAGTCAAAAATGCAATTGTCACCCCTTTTACTTACATACCTCTTTTGCAT 435

QY      42 -----LysAsnProArgLysPheProThrValGlyLysThrIle 54
DB      434 GTATCTCTTTACCTTTCTACCTGAAGAACCCGGGCAAAATTTCTACTGTGCGCAAGACCT 375

QY      54 uGlnLeuLysSerProThrGlyValagLysThrMetLysSerValValSerAlaAlaVa 74
DB      374 CGAGCTGAATCGCCACAGGAGCTGGCAAGCAATGAAGCGTGATTCGCTCTGT 315

QY      74 lGlnGlnLysLysArgMetSerAlaAlaAlaProProSerLysGlnThrTrpArgVa 94
DB      314 GCAGCAAAAGAGCGCATGCTCTGCAGCAGCGGCTCCCTCCACGACATGGCGAGT 255

QY      94 lThrAlaProSerArgProAlaAlaTrpAlaHisProProGlnAlaProLeuValG1 114
DB      254 GACTGTCTCTTCGCTCCGCTCCGCTGCATGGGCACATCCACCTCCACAGGCTCCTCTTGTGCA 195

QY      114 uLysAsnValTyrLysThrProGlnGluProValTyrIleSerProGlnProArgSe 134
DB      194 GAAGAATGATATCAAGACTCCCAAGAAGAGCCGCTATACATATCACCAGGCTCGCAG 135

QY      134 rLeuLysGluAsnLysSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLe 154

```

```

Db      134 TCTTAAGAAATCTAAGCCCTATGACGCCAGAAACCTACTCGAGTGATTGACAAATCT 75
QY      154 uArgPheThrProLeuThrGluThrArgGlyGlyGlnAlaThrIlePheProAspAs 174
Db      74 GCGATTTCACACTCTTCACCGAAACCGG-GGCAAGAGCAGCTACCATTTTCCCGACAA 16
QY      174 nLeuAlaAlaTrp 178
Db      15 TCTGGCAGCCTGG 3

RESULT 6
LOCUS      BC026699                      3130 bp      mRNA      HTC 20-SEP-2002
DEFINITION Mus musculus, clone IMAGE:3994865, mRNA.
ACCESSION  BC026699
VERSION     BC026699.1 GI:19934325
KEYWORDS   HTC.
SOURCE      Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 3130)
AUTHORS    Strausberg,R.
TITLE      Direct Submission
JOURNAL    Submitted (04-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
            Email: cgabbs-remail.nih.gov
            Tissue procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 40 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: no polyA-tail.

FEATURES
            source
            Location/Qualifiers
              1..3130
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="CZECH II"
                /db_xref="taxon:10090"
                /clone="IMAGE:3994865"
                /tissue_type="Mammary tumor metastasized to lung. Tumor
                arose spontaneously from a senescent normal mammary
                (clonal) outgrowth infected with the virus MMV."
                /clone_lib="NCI CGAP_lu29"
                /lab_host="DH10B"
                /note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.:      2,826-50      Length:      3130
Score:          723.50      Matches:    235
Percent Similarity: 43.69%      Conservative: 177
Best Local Similarity: 24.92%      Mismatches: 270
Query Match:    7.60%      Indels:    261
DB:            11          Gaps:        27

```

US-09-914-698-1 (1-1861) x BC026699 (1-3130)

Qy	878	LysPheHisAlaAlaThrValLeuGln-----LysTrpTrpArgArgHisTrp	894	1208	-----AsnTyrLeuHisLeuGlnThrThrThrLysArg	1211
Db	369	AAAGCTGAAATCCTCATCCCTTGTAAATCCAGTTTATGTTTCAAGAGGTGGAAGCGCAAG	428	1374	GCTGCCATTCTCTACAGTCTTATTTTCAGAAATGAGAACAGCGGCGAGCTACCTGAAG	1433
Qy	895	LeuHis-----ValValIleGlnArgArgIleArgHisLysGluLeu	908	1219	-----IleGlnIleLysPheArgAlaLysArgGluMetLysLys	1231
Db	429	CTGCAGTTACAAACCAAAGCTGCAGTACATTCGACGAGCGTTCATGAGAAATGGCATCTC	488	1434	ATGTGCAAGCGCGCTCTTGTTCATTCAGAGTTTCTACTGTGCTTACAGAGCACAATTCAGT	1493
Qy	909	MetArg-----ArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHis---	925	1232	GlnArgAlaGluPheLeuGlnLysLysValThrLeuValValGln-----	1247
Db	489	AGGAAACAAATACGAGAAAGATCCGCTCTGTGATACAGTCGTGTACAGATCCACAGG	548	1494	CAAGGAAGAACTTCTTTGCGAGTCAAGAGGAGCAGCTATTTTGTTGCGCAGCTTACAGA	1553
Qy	926	GlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeu	945	1248	-----LysArgArgArgAlaLeuLeuGlnMet-----	1256
Db	549	GAATTGGCAAAATACATTCATATTT-----AGGCTTGTGTTATGTCATC	593	1554	GGTGCAAAAGTACGCGCAAAATCAAGCAGCAGCTTACGCGTCTGTTCAATTCAGAGG	1613
Qy	946	GlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle	965	1257	-----ArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrIleLys	1272
Db	594	CAGAAAGAGTTTCGTCTTCAAGCCCAAAAGTTATATAGAGAAAGAAAGATGCCATC	653	1614	GTTTTATAGAGCCACAGTCAGAGGATGAATACCACTATGCTTCAGTCGCTGTGAAG	1673
Qy	966	IleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPhe	985	1273	LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly	1292
Db	654	CTGACTCTCCAGACACATTCAGGGCACGTCAGAAGGGGAAGTAGCACACGCTGACTAT	713	1674	ATTTCAGAGATGTACAGGCACACAAAGGTTGCTTATGACATGAGAATTCAATTTCTAAAG	1733
Qy	986	ValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAla	1005	1293	ThrGlnAlaAlaValSerCysLeuGlnMetHisTrpArgAsn-----	1306
Db	714	TTGCAGAAACGAGCTGCCACCATTCGACTCGAGCTCGGTTCCGGGG-----	761	1734	ACTAGAGAGCTGTGTGTTGCCCTCCAGTCTGCCTACCGTGGGTGGCAGGTTCCGACGACG	1793
Qy	1006	LysLysLeuLeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAla	1025	1306	-----CTCAGGAGGCAACATGAGGCTGCAGTGAAGATCCAGTCTACGTTTAGATGGCCGTGGCC	1853
Db	762	-----ATGAAGGCTCGCATTCATCAGGTTG-----CAGATTGGA	797	1307	-----HisLeuLeuArg-----	1310
Qy	1026	AlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHis	1045	1854	CAGCAACAGTATAAACTACTAAAGCAGCTGCAGCAGCTCATCCAGCAGCAGGTGAGGCA	1913
Db	798	GCTGCTCTGTTCTTTCAGTCTCTATGAGAAATGAGACAG-----GAAAGAGTT	845	1311	-----LysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaAlaIleThr	1326
Qy	1046	GluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgMetArgSerLysTrp	1065	1914	CGGGCTGACGGAAGAGGCGACCTTGGCGTACATCCAGTCCGTATCGCTCAGCTCTGGTG	1973
Db	846	AGATTTTAAACCTTAAAGAGATGGTTATTAAAGTTGCAAGCACACATAAGAAA-----	899	1327	LeuGlnArgArgTyrArgAlaArgLeu-----	1335
Qy	1066	SerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGln	1085	1974	TTTCAGGCTGCATGGAAGGAAAGATGCTGAGAAGACAGATTGCCCGCAGCATCAATGC	2033
Db	900	-----TACCAACAACCTACAGAAATATAAGAAATAAGAAAGCAGCCATTACAAATTCA	953	1336	-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu	1347
Qy	1086	GlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArg	1105	2034	GCTGCCTCATACAGTCATATTACAGGATGCGCATTCAGCAGGAGTGGAGCATCATG	2093
Db	954	ACTCATTCGAGCTTCCATTTTCAGCCAGGAGGTTCTAGCATCTTATCAGAAACAAGG	1013	1348	LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLeuAlaMetGlnLysGln	1367
Qy	1106	SerSerValLeuLysValGln-----AlaTyrArg	1115	2094	AAAAACGCTGCCCTTCAGATTTCAGCTGTGCTATAGAGCTTACAAAGTTGGAAGAAGACAG	2153
Db	1014	TCTTCAGTCATTCTTCTACAGTCTGCATGTAGAGGGATGCAAGCCAGAAAGCATTTCCG	1073	1368	ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArgGly	1387
Qy	1116	LysAlaThrIleGlnMetArgIleAspArgAsnHisTyr-----TyrSerLeuArgLys	1133	2154	CGCCACTTATCTTGAACAACAAAGCAGCTGTAGTAACCTTTACAGTCAGCTTACCGTGT	2213
Db	1074	CATGCTTACATCCCGTATATAAAGATTCACTCATATTATCGGCTTATATATGATGGAAG	1133	1388	AsnLeuGluMetArgLysGlnIle-----	1395
Qy	1134	-----AsnValIleCysLeuLeuGlnArgLeuArgAlaIleMetLysMetArgGlu	1150	2214	-----ATCAAGAGTCAGAAACGGGTGCTGATGCCAACAGCAGCGTAATATACAGTCT	2270
Db	1134	ACATTTTCAGAACTTCAAAAATGCCACAATAAAGCTGCAGTCAATTGTCAAAATGAACAA	1193	1396	-----GluValTyrGlnLysGlnArgGlnAlaValIleArg	1407
Qy	1151	GlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArg	1170	2271	AAATTTAGAGCCCTACAGAACACAAAGAAGTACACGACTATAGAACTTCAGCTATTGTA	2330
Db	1194	TCACGAAGCAGATTTTACAAATAAAGACGCTGCACTGTTTATCCAGCGGTGGTACCGT	1253	1408	LeuGlnLysTrpTrpArgSerIleArg-----	1416
Qy	1171	MetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle	1190	2331	ATTTCAGAGTGTACCGGAATATTAATATACACCCCAACATCAGGAATACTTTAAC	2390
Db	1254	TCCCAAAACTGGCTCCACAAAAGAGAAAGACTACATACAGGTGGTGTGNAATCTGTATC	1313	1416	-----TTAAGGAGGCGCAGTGTAAAGTCCAGGCTGCTGATAGAGGCATTAGAGTGAGGAGACGC	2450
Qy	1191	AsnValGlnArgArgTrpArgAlaThrLeu-----GlnMetArgArgGluArgLys	1207			
Db	1314	AAACTGCAGTCACACTTTAGAGGATCCCTGGTCAGAAAGCAGTTGAGGTTGCAACGTAAA	1373			

QY 1417 -----AspMetArgLeuCys 1421  
 Db 2451 ATTGACGACATGACATGCGCGCCACACTTATTAAAGCCATGTTTAAAGTGGTTCAGTCT 2510  
 QY 1422 LysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAla 1441  
 Db 2511 AGAGTAAGGTACTCTCAAGATGAGACAGCAGCTCTTATTAATTCAGGTGAGATACAGCG 2570  
 QY 1442 ThrValGlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeu 1461  
 Db 2571 TATTATCTAGTGAATTCAGTCAGTGAAGTACCTGAGGACTTTGAAGGCTATTAAACC 2630  
 QY 1462 MetGlnAlaPheIleArg----- 1467  
 Db 2631 CTTCAAGCTGGAGTCAGAGGACCAAGAGTGGAGCGGACTGTGAAGAAAGTGCATTTCGCA 2690  
 QY 1468 AlaThrLeuLeuMet-----ArgGlnGlnArgArgGluPheGluMetLys 1482  
 Db 2691 GCACACTCATTCAGTCACATTTTAGGGGACACAGCGCAGCAAACTTACTTTCATCGGTG 2750  
 QY 1483 ArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAla 1502  
 Db 2751 AGAAGGCGGCAACCATGGTCAGCAGAGATACCGAGCAGTGAAGAGGAAAGCGCTGAG 2810  
 QY 1503 ArgGlnAspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPheArgAla 1522  
 Db 2811 TTTCAGAGGTACAGCAGACTAAGCGCTTCTGTACTCTCATCCAGGCTGCCTTCAGGGGA 2870  
 QY 1523 AsnArgSerMetLysGln----- 1528  
 Db 2871 CTGAGAACTAGGAGACACTTAAAGCTATGCACCTGGCTGCTACTCTCATTCAGAGGAGA 2930  
 QY 1529 -----AlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
 Db 2931 TTCAGGACTTTCGCATGAGAGAAAGTTCTTTCCTCCAGAAACTGCCATTGGATT 2990  
 QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
 Db 2991 CAGAGCAGTACCGTGCA---AGACTTATGCGCAAGTATTCGCGACACAGTACTGTG 3047  
 QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
 Db 3048 GAAAGGCGAGTTATAAGATCCAGTCATCATCAGAGGATGGTGGTGAAGGAAAGGCTG 3107  
 QY 1584 GlnAlaLeu 1586  
 Db 3108 CAAAGATG 3116  
 RESULT 7  
 AK052326  
 LOCUS  
 DEFINITION  
 Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched library, clone:D330028K02 product:in hypothet  
 Transglutaminase/calponin-homology domain, CH-domain structure containing protein, full insert sequence.  
 AK052326  
 VERSION  
 AK052326.1 GI:26095125  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 JOURNAL  
 99279253  
 MEDLINE  
 10349636  
 PUBLISHED  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20493374  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 3983)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, R., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/.  
 Location/Qualifiers  
 1. .3983  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:D330028K02"  
 /db\_xref="MGI:2421386"  
 /db\_xref="taxon:10090"  
 /clone="D330028K02"  
 /tissue type="heart"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 dev\_stage="13 days embryo"  
 1. .3983  
 /note="Hypothetical Transglutaminase/Calponin-homology domain, CH-domain structure containing protein (SCOP|49309, SCOP|47576, evidence: SCOP)"  
 misc\_feature  
 1. .3983  
 /note="Hypothetical Transglutaminase/Calponin-homology domain, CH-domain structure containing protein (SCOP|49309, SCOP|47576, evidence: SCOP)"  
 ORIGIN

Alignment Scores:		1,71e-49	Length:	3983
Pred. No.:		716.50	Matches:	268
Score:		43.11%	Conservative:	167
Percent Similarity:		26.56%	Mismatches:	388
Best Local Similarity:		7.53%	Indels:	186
Query Match:		11	Gaps:	39
DB:				
US-09-914-698-1 (1-1861) x AK052326 (1-3983)				
QY	1	MetGluLeuValTrpSerProValLeuGluValAlaCysLysGluThrLeuLeuLe 20		
DB	452	ATTTCGTTACCTGACCGCCTGCGAGAGGGGAGTGGAGGAGATTGTACATTCTT 511		
QY	21	AspAsnArgAsnPheArgLysGluValMetIleIleLeuLysSerLysSerAsnGlnPro 40		
DB	512	GTAAAT---GATTCTCTGAAGCACCAGGCTATATTA---GGAAATGCAGAGAGCCT 565		
QY	41	ValLysAsnProArgLysPheProThrValGlyLysThrLeuGlnLeuLysSerProThr 60		
DB	566	AAGAAGAAAGAGAGTCTTTGGAATACCAGTAAG-----AAGATTCCAGCC 613		
QY	61	GlyAlaGlyLysThrMetLysSerVal-----ValSerAlaAlaVal 74		
DB	614	TCCTCAAACATACAAAAGGACTTCCAAAACCAACATTTTAATGAATCATTTACTATT 673		
QY	75	GlnGlnLysLysArg-----MetSerAla 82		
DB	674	TCACAAAAGACAGATTAGGAGCCCACTGACGCTTGTGAAATCTGGCTATGAGTAA 733		
QY	83	AlaAlaAlaProProSerLysGlnThrTrpArgValThrAlaProSer---ArgProAla 101		
DB	734	TGCTCTTCCCAACAGAAAC-----AAAGTCCCAACCCATCCATTAGTCTCTATT 784		
QY	102	AlaTrpAlaHisProProGlnAlaProLeuVal-----GluLysAsnValTyLys 119		
DB	785	AGAGAATGCCAGAGTGAACCTGCTGCCACTGTTTAAACGCGAGTCCACTGCCTATTCA 844		
QY	120	ThrProGlnGluProValTyLLeSerProGlnProArgSerLeuLys---GluAsn 138		
DB	845	TCTCTTCATGAA-----TCTGAAATAACACAAATTTAAAGATACAGAT 889		
QY	139	LeuSerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPhe---Thr 157		
DB	890	GCCAGCATTTCAAACTTTTGAATTTAATAGGAAGTGCCTGCAATGAACTTTATTATAT 949		
QY	158	ProLeuThr-----GluThrArgGlyLysGlyGlnAlaThrIlePheProAspAsn 174		
DB	950	CCCATTAGTGTCTGCACCAAGAGTGAAGGGGATAGGAACTCAGCTTGCCCCAAAC--- 1006		
QY	175	LeuAlaAlaTrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAsp--- 193		
DB	1007	---TGTTCTTCACCTTTGAATAGTACACAGCCCAATACATCTTCTAAGTCCAGATTCT 1063		
QY	194	---MetArgProArgGlleThrProAspAspLeuGluAsp---GlnProAlaThrAsn 211		
DB	1064	TTTGTAAACACAGGTATACGTCTGATAATGACCTGAAATCAATGAAATGTTTTATCA 1123		
QY	212	LysThrPheAspValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp--- 230		
DB	1124	GACACATTTAGGAAGATCCCGCAGAGTCTGTATGCTGGAATCCCAACTGTCCATGAA 1183		
QY	231	---CysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal 249		
DB	1184	GTTTGCCAAACATC---TTAAGCCAGATCTTTTTTAATGATATATGACTTAA 1240		
QY	250	HisAlaThrHisThrArgAlaLeuAlaCysIleHisGluGluGluGlyProSerProPro 269		
DB	1241	AAAGGGCTAAACTCAAGTCGGTC-----AATCCCGTT 1273		
QY	270	ArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuValGlySer 289		
DB		-----		
Db	1274	TTGTCGCCAACACAAATTT-----GTAAAGGATAGCATGGGACACAGTAGGTCTAG 1321		
QY	290	ProLeuArgLysTySerGluSerMetLysAspLeuSerLeu-----Leu 304		
Db	1322	CAAACGGGTAAAGTCAAAATGAAGCTTCTCAGGATTGGAGATAAATGAAGGTTGGCGTAT 1381		
QY	305	SerProGlnThrLysTyTrpAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIle 324		
Db	1382	ACCCCGAGTGTGAGCATGCA-----CAGACTCCAGCAGCAGATCTGAAAAACAA 1432		
QY	325	ArgSerIleGlu----- 328		
Db	1433	AACCTGTGAAGTGAAGCACAATCGTATGATTTCAAGCAAAAGCCCAAGATTCT 1492		
QY	329	-----GlnAsnArgTyTrpGlnGluGlnGlnIleGlnIleLysAlaLysAspLeuAsn 346		
Db	1493	GAGTTTCAGGATGCTTTTGTCCACAGAGTAACACCCGACACAGAGACCCCAATTTCTT 1552		
QY	347	SerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSer 366		
Db	1553	TCTGCCACTGTTACGAAAGAAACCCACCAATGCCAGAGAAAGCTA-----CCT 1603		
QY	367	GluIleLeuAlaGlnSerSerArgPheAsnLeuHis----- 378		
Db	1604	GAGATTAAACAGCCAGACGCAAAAGATGTCTCGAAGTCTAGTAGTGACGTGGAATA 1663		
QY	379	GluValGlyArgLysSerValLysGly-----SerProValLysAsnPro--- 393		
Db	1664	GAAGTGGTAGCCTGAGAGAGAAAGGTTTCATTCTCTCTCCGTTGGAGCCAGGA 1723		
QY	394	-----His 394		
Db	1724	GTGAGTAAAGCTCTGAGCTACAGGATGAAGTAACGCTGCAACTGTTGCTGCTCGG 1783		
QY	395	LysArgArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyArg 414		
Db	1784	AAAAGAAAGAGCCACCGA---ACGTGGGAGATGCA-----AATGAAAAGTAGCAGCT 1834		
QY	415	AsnGluThrValAlaIleSerProLysLys-----GlnArgVal 428		
Db	1835	GAAGAATGGATGGACATGTGTGAAGTCAACAGAAATCCATTTTCTCTCTGGAATCTACA 1894		
QY	429	GluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSer 448		
Db	1895	CCCTCAACTGTTCAGAAACCAACAAAGAAAGAGGCGATCACTCAAAGCGGATAAGCAGC 1954		
QY	449	AlaHisAlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeu 468		
Db	1955	-----CTGGAGCGGTGAGGCTGAGAGAAAATGGATTTCATCAATA 1996		
QY	469	MetLysLysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLys 488		
Db	1997	TTAAACACCTCTTTCTAAGACAAAGAAAAGAAAGATCCATT-----GTTGCT 2047		
QY	489	LeuTyAspSerGluLeuTyMetGlnThrCysIleAsnProAspProPheAlaAlaThr 508		
Db	2048	GTAGCACAGTCTCATTTG-----ACCTTTATAAGCCATTAAAGACGCTATTTCCT 2098		
QY	509	ThrThrIleAspProPheLeuAlaSerThrMetTyLeuAspGluGlnAlaValAspArg 528		
Db	2099	AGACCCCAATGCGTTTGTGTCAGAAACACATGTTATATGACGACGCTGGAAGAAAAG 2158		
QY	529	HisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAsp 548		
Db	2159	CAGGAACAGGGCTTCACTTGGTGGCTAAATATATATTAACA---CCTGATGATCTTCACT 2215		
QY	549	AlaAspLeuAsn---AsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLys 567		
Db	2216	GTAAACCAATGTTCTAAAGTGAACGCTGCCAGCCTTGTGTTGGAGCGGGAAGCCAA 2275		
QY	568	GluLeuVal-----ValAlaProThrLysGluGlnGlnSerMetAsn---TyLeu 583		
Db	2276	CATAAATAAGTGTCCCGAAAGACCTACTATAAGAGGAAGTGTTCTCTCAGAGCTTATACA 2335		



```
Qy 584 ThrLysTyrArgLeuGluThrLeuArgLysAlaIaValGluLeuPhePheSerGluGln 603
Db 2336 GCCAGCTCGAGCTGAATAGATTACCGGTACAGCTTGCTCTGTTTACTTCTGAATA 2395
Qy 604 MetArgLeuProCysSerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArg 623
Db 2396 ATGGTTAAAGCTATTAAAAAGGTTGAAATCGAAATGAAGTTGGCGGCTACTTGTCTGA 2455
Qy 624 SerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeuLeu 643
Db 2456 AAAGACCCACACCTCTGGAAGATATCGAACACGCCCAAAAGTCTGAATTGGCTCTGTG 2515
Qy 644 CysPheAsnProLeuTriPheArgLeuGlyLeuGluValValPheGlyGluLysIleGln 663
Db 2516 TCATATATCTCTGTGGCTCGAATTGGTCTGGAGACAGTTTTTGGAGAACTGATCCCT 2575
Qy 664 MetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArg 683
Db 2576 CTGGCAGACAAACAGTGACGTACCGGATTGGCGATGTTCAATCTGAATCGCTTACTTTGG 2635
Qy 684 AsnLys-----CysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGlu--- 700
Db 2636 AATCTGACATAGCAGCTGAAATATAGACACCCCACTGTTCTCTCTCTGTTTGGGATGGT 2695
Qy 701 TyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeu 720
Db 2696 CATGAGCAGCTGCTCTAAGTTACATGTAAGAAATGCTGTTGTTGATCTGTTCTCTT 2755
Qy 721 AspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSer 740
Db 2756 GATCATGCTAAATTTCCAGACTTATTCATGATCATGATCTTGTTCTTCTGTAAGATGCT 2815
Qy 741 ProHisLysGluThrLysAspIleLeuArgPheSerSerGluLeuLeuAlaAsnIle 760
Db 2816 GAATTCAGGCTAGCAAGAGCTCTTCTAGCTTTTCCAGGATTTCTCAGTGGTGGA 2875
Qy 761 GlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPhe 780
Db 2876 GGTGACCTTCTGTCACCTAGCTTCTGTTGGGTACCTGTGAGCAGCAGTTCAGACACA 2935
Qy 781 LeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArg 800
Db 2936 CTTGATGAATTTGATTTGCTTTACAAATCTTGCAATGTCAGTGGACTTGCACTGGGTTCG 2995
Qy 801 LeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal 820
Db 2996 CTTGTGCAACTGTGGAGCTTCTCACACAGAACTGGAATCTGTCCGCAAACTCAGGATC 3055
Qy 821 ProAlaIleSerArgGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGly 840
Db 3056 CCTGCATACTCGAGTTCAAGATGACCAATGTTGATCTGCTCTGCTTCAGTTCAG 3115
Qy 841 Glu-AlaAsnPheGln----- 845
Db 3116 TCTCGGGAGTTCCACTGACTGACGAGCATGTGTAACGAGCGTCAGAAATAAGCTTCGTG 3175
Qy 846 ----- 847
Db 3176 TCCTTGTGAAGCTTGATGTGTGACGTACCGTCTCTAACGGACTGCTTTCCTTTGTAGG 3235
Qy 847 yGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLe 867
Db 3236 GAGCGCTATCTGCTTAGAGATGTTGTGACAGGCACAGAGAGAAACCCCTTGGTTCCT 3295
Qy 867 uTrpGlnLeuIleTyrLysPheArg 875
Db 3296 TTGGAAATAGCACGTGGCTTTCAG 3320
RESULT 8
Bg636223
LOCUS BG636223
DEFINITION SD13838.5prime SD Drosophila melanogaster Schneider L2 cell culture
```

pot2 Drosophila melanogaster cDNA clone SD13838 5 similar to asp:  
Fban0006875 'cytoskeletal structural protein' located on: 3R  
96223-96A23; 04/13/2001, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BG636223.1 GI:13763760  
EST.  
BG636223.1  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 589)

REFERENCE

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.

TITLE

BDGP/HMMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd., Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

Based upon one or more reads of this clone where vector sequence  
was present at both ends, this clone has been determined to contain  
contain a cDNA insert on the order of 600-1000 bases. hit genomic  
AE003749: arm:3R [20443564,20672042] estimated-cyto:96A20-96B10:  
04/13/2001

Plate: SD.138 row: D column: 2

High quality sequence stop: 547.

FEATURES

source

1..589  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="SD13838"  
/lab\_host="DH5-alpha"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture pot2"  
/note="vector: pot2; site 1: EcoRI; site 2: XhoI; Sized  
fractionated cDNAs were directly ligated into pot2.  
Plasmid cDNA library."

ORIGIN

Alignment Scores:  
Pred. No.: 5,96e-45 Length: 589  
Score: 647.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.80% Indels: 0  
DB: .12 Gaps: 0

US-09-914-698-1 (1-1861) x BG636223 (1-589)

Qy 1743 ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPhe 1762

Db 1 CGCTGGTGGCAACAGACAGTGGATATTCACACTTTGTGCACCTCATTTGGTATTC 60

Qy 1763 AlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIle 1782

Db 61 GCCCACTGCTCTAAAGACGAAGATCATTCACGACTATATGACCAACCCAGAGGCCATT 120

Qy 1783 TyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAla 1802

Db 121 TACATGGTGGCGGAAACTAAGAGCTGTGGCCCGCAGGAAAGATGAAGCAAAATGCA 180

Qy 1803 ArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrPro 1822

Db 181 CGCAAGCGCGCGCAATGACAAAGTGGCGGTATTAAGAGCCAAAGATAAACTTCAGCGCG 240

Qy 1823 CysSerLeuProSerLeuGluProAspPheGlyIleLeuArgTyrSerProTyrThrPhe 1842

Db 241 TGTTCCTGCGCAGCTGGAGCGGACTTCGGAATCATCCGCTACAGTCCCTCACAGTTT 300



Db 277 TTACAGAAATGCACAGGCTGCAGCTCTCATCCAGGCTACTTTTAGA----- 324  
 QY 1094 MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAla 1113  
 Db 325 -----ATGCACAGAACTTACGTTGTCATTTCCACATT 354  
 QY 1114 TyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLys 1133  
 Db 355 TGGAAATGTCCTCGATTAGA----- 375  
 QY 1134 AsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGlu 1153  
 Db 376 -----ATTACAGCAGTGTATCGAACATACAGAACTATAAAATTACAAAAAGAA 423  
 QY 1154 AsnTyrLeuArg-----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMet 1171  
 Db 424 AAACCTTATCAGAGAGAACACACAGTGTCTGTCTTAATATCAAGACGATACAGGATG 483  
 QY 1172 ArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsn 1191  
 Db 484 TACAGGCAG-----CGTTGCTTCTATCAA 507  
 QY 1192 ValGlnArgArgTyrArgAlaThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHis 1211  
 Db 508 -----CAGCGTGGTGGCAGCA----- 525  
 QY 1212 LeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLys 1231  
 Db 526 -----AAGTATACAAAAAATACAGAGCA-----AATAAA 558  
 QY 1232 GlnArgAlaGluPheLeuGlnLeu-----LysLysValThrLeuValValGlnLysArg 1250  
 Db 559 AGACGCGAAGATCTCTCTATGTGTGTAAAGGAAGAACACCTCTTCTCAGATGCTATT 618  
 QY 1251 ArgAlaLeuLeuGlnMetArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThr 1270  
 Db 619 CAGGGTTTAAACACTGCCAAAACAGTCGCGCAGCAGCAT-----GGGGCGCGCT 666  
 QY 1271 IleLysLeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyr 1290  
 Db 667 ATGATTACAGAAGACATTTTAGAGCCTTTAAAGCAAGAGGCTGATGGAGCCGAG--- 723  
 QY 1291 ArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArg 1310  
 Db 724 AGAGCTTCCAGCTCGA-----TGTAGA 747  
 QY 1311 LysArgGluArgAsnSerPheLeuGlnLeuArgGlnAlaIleThrLeuGln----- 1328  
 Db 748 AAATATAAAAGCTAAGAAATACTTATCCAAAGTGGAAAGCTGCTGCAGGATTCAAGCTTG 807  
 QY 1329 ---ArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
 Db 808 TATAGCGCTTGAGAGCGGAC-----AAGAAATATCTAATCTTA 846  
 QY 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
 Db 847 TTGAAAGCTTTAATCATTTAAGAGTTACCTCTCTGCACAGCTGCGCAGAGGAGG--- 903  
 QY 1368 ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGly 1387  
 Db 904 -----TTCTTGAAGATGAGGCGCGCAGCTATTATAAATCAAGAAATGAGAGCT 954  
 QY 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
 Db 955 ACACCTTCTGTAAGAGAGCTCGTGAGAAATTTAAAGAACATCGAGCTGCTTGTGTG--- 1011  
 QY 1408 LeuGlnLysTyrTyrArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArgArg 1427  
 Db 1012 ATTCAGCCCACTTTAGGGCTATCAG-----GCNAGGCACTCTCTCTGCAG 1059  
 QY 1428 IleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArgArg 1447  
 Db 1060 CAGAGATCTGCTGTTTAAATCATACAGACATGTACGAGCCATGCTGGTGCAGAGCAG 1119

QY 1448 GlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArg 1467  
 Db 1120 GAAAGTAAGATATATAAATTGAAAGTCTACGGTTGTTGTACAGCCCTTGGTGGCT 1179  
 QY 1468 AlaThrLeuLeuMetArgGlnArgGluPheGluMetLysArgArg----- 1484  
 Db 1180 GGC---TGGTTAGTAAGAAAGGTTTCAGAGCAGAAAGCCAAACTCGGCTTTTCCAC 1236  
 QY 1485 -----AlaValValIleGlnArgArgPheArgAla 1495  
 Db 1237 TTTACAGCAGCTGCCCTACTGTGCATGTGTCTTAAATCCACAGCTTATCGGCTC 1296  
 QY 1496 ArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeu 1515  
 Db 1297 CAGGTGACCTTGAGAACGCTAAGAACAC-----ATGGATTCACTATCTTC 1344  
 QY 1516 ValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArg-GlnGluPheValGln 1535  
 Db 1345 ATTCAGAGATGCTTTCGAAAAAAGATTACAGAAAGAGGTTTCATTGACAGTACCACAA 1404  
 QY 1535 nLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGln 1555  
 Db 1405 GATC-----CTAAGCACTCGACGCTGAAGCTCATGCTGTTGGCTCCA 1446  
 QY 1555 uGlnArgAsnCysPheGlnLeuArgCysSerMetProGlyPheGlnAlaArgAlaArg 1575  
 Db 1447 CGAGACAGGCTGCTATCCGTGATA-----CAGAAGCGGTACG 1485  
 QY 1575 gGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIle 1595  
 Db 1486 CGCTTTCTCTCTCGCAGAGACAGGAAAAAGATCACTAGC----- 1525  
 QY 1595 eArgGlnLysArgAlaAlaLysValIleGlnArgTyrTyrArgGlyTyrLeuIleArgArg 1615  
 Db 1526 -----TGCGCCACTAGAAATTCAGGCATTCAGAGAGCTATCTTGGAGGAA 1572  
 QY 1615 gArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGln 1635  
 Db 1573 GAAAAATGACACACACAGAAATCAAAAGCTATCCGAGAACCTTAGAGCTGTGAGCAGCAG 1632  
 QY 1635 uAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGln 1655  
 Db 1633 CGTGAGAGAGAGAACAAAGCTGTACAGAAAGAACTGAGCGAGCGCTTCACCTGCTCAC 1692  
 QY 1655 yArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgTh 1675  
 Db 1693 CTACAAACACTTGTGCGCCATTCTGGATGCGTTGAAGCACCCTGGAGGTCGTTACCAGACT 1752  
 QY 1675 rValProHisLeuLeuMetTyrCysSerGluPheMetSer-----ThrPh 1690  
 Db 1753 CTCTCCA-----CTTGTGTGAGAAATAGCTGAAGCGGAGCGGTATCTAC 1800  
 QY 1690 eCysTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArg 1710  
 Db 1801 CATCTTGTGTGATCCGAAGCTGAATCGAGTGTTCTTGTATGGAAGTCGTTGGATA 1860  
 QY 1710 gCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
 Db 1861 TGCTGTTCAAGTATTGCTGAATGTGGCAAGATATGATAAGACGATCGCGCTGTGTACGA 1920  
 QY 1730 n---GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
 Db 1921 GGCCGAGAACTGCGGGACACACTGCTGGAGTCTCTCGAGGTGACCGAGAGAACCTGG 1980  
 QY 1743 ---ArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTyrVa 1761  
 Db 1981 AGACCGAGTCTCGAGAGAGTGCAGCATTTTACAGAACTTCTGTTTGTAGCAGT 2040  
 QY 1761 lPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAl 1781  
 Db 2041 GCTGTG-----AAGACAGAGCAGTGTGCTTTTGTATGCACAAAGTAGATCTAAAGT 2091

```

QY 1781 alleTyMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAs 1801
Db 2092 CACTGATCGATTATTCGCTCTACAAATTTACAGTTCCTTAAGCAATAAGTAAT----- 2146
QY 1801 nAlaArgLysProProMetThrSerGlyArgTyTrLysSerGlnLysIleAsnPheTh 1821
Db 2147 -----ACACAGGATTGTTTGATAGACAGAGAGCAATTCCTG 2184
QY 1821 r-----ProCysSerLeuProSer-----Ie 1828
Db 2185 TGTAGGCTTTCCTTGT-----ATACCAAGAAAGAACTATGAAGCACCAGGTAGTTTCAAGACT 2241
QY 1828 uGluProAspPheGlyIleLeArgTyTrSerProTyTrThrPheIleSerValTyAl 1848
Db 2242 TAAGCCACAGTGGTGTGGGGGTGTATATGTGGAGAAATACAAACTCCCTGCAAGC 2301
QY 1848 aPheAspThrIleLeuCysLysLeuGlnIle 1858
Db 2302 TATTCAACTGGTAATGATACACTTGGAAAT 2332

RESULT 10
BH374743/c
LOCUS BH374743 485 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-143H23.TF ND-TAM Anopheles gambiae genomic clone
AG-ND-143H23, genomic survey sequence.
ACCESSION BH374743.1 GI:17320885
VERSION BH374743
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 485)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J.,
Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B.,
Gardner,M.J. and Collins,F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
12655398
Other GSSs: AG-ND-143H23.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(IGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..485
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-143H23"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

FEATURES
source
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2,34e-26 Length: 485
Score: 429.00 Matches: 78
Percent Similarity: 71.25% Conservative: 36
Best Local Similarity: 48.75% Mismatches: 46
Query Match: 4.51% Indels: 0
DB: 28 Gaps: 0

US-09-914-698-1 (1-1861) x BH374743 (1-485)
QY 848 GlyAspIleAlaGlnAspIleValAspGlyHisArgGlnLysThrLeuSerLeuLeu 867
Db 482 GGCACATCCCGCCCAAGGACATCTCGACGGGCATCGGAGACAGATGTCCTGCTC 423
QY 868 TrpGlnLeuIleTyTrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln 887
Db 422 TGCAGATTGTGTACAAATTCGTGGCCCAAAATTCATGCGGCCCGCAATCGTGTGCAG 363
QY 888 LysTrpTrpArgArgHisTrpLeuHisValIleGlnArgIleArgHisLysGlu 907
Db 362 CGCTGGTGGCTATGAACCTGGCTGAAGGTGACCATTTCCCGCGGATCGAAGAAACGT 303
QY 908 LeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMet 927
Db 302 GCCTTGGCGCGAGAGCGGCGCCCGCAGCATTCAGGCCCTCTCCGCGGTACTGCGTG 243
QY 928 ArgLysTyTrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947
Db 242 CGCGTTTGTGACGAAGCCCATCGACGGCAGAAAGTGGCTGCGCGGTATCGACGATCAGCG 183
QY 948 PheThrArgArgTyTrLeuAlaGlnLysGlnLeuTyTrGlnSerTyTrHisSerIleIleThr 967
Db 182 TTCTCCAGACGATATCTGCGCGAGAAAGTGGCTGCGCGCGGTTCAGCGCATCGTGC 123
QY 968 IleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGlu 987
Db 122 ATACAGCAATGTTGGCGAAGCGTGGCGCCAAATCGCCAGCGCGGAGCGTTCCTGCTC 63
QY 988 LeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLys 1007
Db 62 TGCAGAAAGTCAGCAATCTGACTGTGCAAAACATCTACGACGCTACGCGCTCGGACGAG 3

RESULT 11
CNS04QCR/c
LOCUS CNS04QCR 1008 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
129E12 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL302436
VERSION AL302436.1 GI:8181338
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Rukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
JOURNAL

```







```
Db 792 AAACAGGAAATGCTGCTCTGTAACGCCAAGAAAAGA-----GCTGTAATCATC 745
QY 1300 LeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGln 1319
Db 744 TTACAAAGCTGCTTTCCGTGA-----ATGAAATGCAGAGTCAATTAAACGA 697
QY 1320 LeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLys 1339
Db 696 AGGCACCAGCCGCCGATGTCATCCAGAGAGTTTATAGACCCGC-----CGT 649
QY 1340 GlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArg 1359
Db 648 GAACGCAAGCAGATTCATCTTTGAGATCTCTGTGTGACCATCCACGGAGATATCGA 589
QY 1360 AlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIle 1379
Db 588 GCCACTGTGGCTCAAAAGAGAGTAAGTAAACTCTACCAACGGATGCCAGAGCAGCAGTC 529
QY 1380 LysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGln 1399
Db 528 CTGCTACAGGCAGCCTTCAGGGT-----CAGCGCGGTGAGAAAGAGTGGCGCGC 478
QY 1400 LysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg 1419
Db 477 TGGCATCGAGCTCAACTGTG---ATCCAGTCTGCATTCAGAACCCACAGCAG----- 427
QY 1420 LeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrp 1439
Db 426 -----CAGATCAAGTTCAGGCCATCGCTCTGCCCTGTGGTCAATTCAGAGATACTAC 373
QY 1440 ArgAlaThrValGlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysVal 1459
Db 372 CGCTCTGCAATCTCCAGAGCGCGGAGAGACACCTTCCTGGAAACGAGACGTTCTGCC 313
QY 1460 ArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPhe 1479
Db 312 ATCGTCTCTTCAGGCAGCTTTCAGAGT-----CAGCGCGTGGCGCGCGAGTGC 265
QY 1480 GluMetLysArgArgAlaAlaValIleGln-----ArgArgPheArgAlaArg 1496
Db 264 AACAACTGCACAGAGTCGCATGCTATCCAGCCCACTTCAGAGATTTAAGCTCAG 205
QY 1497 CysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuVal 1516
Db 204 TCTGCATTTAGGAGCAGCGCCAG-----GCTGCTGTGTCTCTC 166
QY 1517 GlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeu 1536
Db 165 CAGCAGAGGTTCCAGAGCCCAAGACAGAGAATGTGGAGGTTAAACATTTACCAACAGCTC 106
QY 1537 ArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGluGln 1556
Db 105 AGAAATGCTGCTGTCGTGTGAGGCTGCATATCGTGAATGAAA----- 61
QY 1557 ArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGly 1576
Db 60 -----TGCAGAAATATCATCAAA----- 43
QY 1577 PheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArg 1596
Db 42 -----CGG 40
QY 1597 GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArg 1609
Db 39 AGGCATCGGGCTGCAAGTGTTATTTCAGCGGGCTTTTAGA 1
```

Search completed: September 15, 2004, 09:09:40  
Job time : 9641 secs



1	313	3.3	6856	4	US-09-566-921-42	Sequence 42, Appl
2	294	3.1	9551	1	US-08-056-200-93	Sequence 93, Appl
3	294	3.1	9551	2	US-08-800-644-93	Sequence 93, Appl
4	277	2.9	5467	4	US-09-576-594-735	Sequence 735, Appl
5	273.5	2.9	8503	4	US-09-620-312D-130	Sequence 130, Appl
6	273.5	2.9	580073	4	US-08-545-528D-1	Sequence 1, Appl
7	271.5	2.9	8257	4	US-09-595-684B-30	Sequence 30, Appl
8	268	2.8	8789	1	US-08-328-254-5	Sequence 5, Appl
9	259	2.7	10136	1	US-08-353-700-2	Sequence 2, Appl
10	259	2.7	10136	5	PCT-US95-16216-2	Sequence 2, Appl
c 11	257	2.7	2214	3	US-08-864-038A-1	Sequence 1, Appl
c 12	257	2.7	3331	3	US-08-864-038A-2	Sequence 2, Appl





QY 1514 --lleuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluP 1533  
 Db 4638 CAGTCACATTCACGAGAAAGAAAGATTTCAGAGGATCGTGGAAATACAGAGGAGG 4697  
 QY 1533 heValGlnLeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuM 1553  
 Db 4698 ATGAGCAAAACACTGTTAAGAACCTGATTCGAACTGAAGCCACGCGTGGTAGCAGTCA 4757  
 QY 1553 etIleGlnGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaA 1573  
 Db 4758 ATTG-----ATTCCAGGATTACCGGCAT 4781  
 QY 1573 rgAlaArgGlyPheMetAlaArg-----LysArgPheGlnA 1585  
 Db 4782 ATATCTGTCATGTGTTCGATCGCTGACTACCTGATGATGATGATGATGATGATGATGAT 4841  
 QY 1585 laLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValileG 1605  
 Db 4842 CGTTGCTAACATCAACAATTAACAGCATC-----AAAAAGATTGA 4883  
 QY 1605 lnArg-----TyrTrpArgGlyTyrLeuIleArg----- 1614  
 Db 4884 AGAAAAGAGTGATGATTTTGAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4943  
 QY 1615 -----ArgArgG 1617  
 Db 4944 TGCACTGCTTGAACAGTACAGTGGAGAGAGGGCTTTATGAGAGCAACACATCTCGCC 5003  
 QY 1617 lnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaL 1637  
 Db 5004 AGAATGAACACTCGCTCAACCAATTTT-----GACCTGGCTGAGTATCGGCAG----- 5050  
 QY 1637 ysAlaValAsnSerValArgCysLysVal-----GlnGluAlaValArgPheLeuArgGly- 1655  
 Db 5051 --GTGCTGATGATCTTGGCAATTCAGATTCACAGCAGCTCGTGGCGGGTGTAGAGAAC 5108  
 QY 1656 -----ArgPheIleAlaSerAspAlaLeu----- 1663  
 Db 5109 TCCTTCAGCAATGATGTCTCAGCATGTCTGAAACATGAAACATGAAACATGAAACATGAA 5168  
 QY 1664 -----AlaValLeuSerGlnLeuAspArgLeuSerArgT 1675  
 Db 5169 GGGTGAAGCCACAGGCTTCAGAAAGCGAACCTCCAGTATCGCGCAGAGGCGACCTACA 5228  
 QY 1675 hrValProHisLeuLeuMetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleM 1695  
 Db 5229 CACTGGAATCCCTCCGCGAGCTCAACTCCTTCCACTCGGTCATGTGTGACGATGGCA 5288  
 QY 1695 etAlaGlnAlaIleArgSerGluValAspLysGlnLeuIleGluArgCysSerArgIleI 1715  
 Db 5289 TGGACCTGAACTGATCAAGCAGGTGTGTCAGCAGATGTTCTACATCATAGGGGCCATCA 5348  
 QY 1715 leLeuAsn-----LeuAlaA 1720  
 Db 5349 CCCTGAACACCTCTCTCTCGGAAGGACATGTGCTCTCGAGTAAAGGCGATGCGATCA 5408  
 QY 1720 rgTyrAsnSerThrVal-----AsnThrPheGlnGluGlyG 1733  
 Db 5409 GGTACAAATGATGATCAACTGAAAGATGGTGGTGGTGAAGAATCTGATGAATAGTGGGG 5468  
 QY 1733 ly-----LeuValThrIleAlaGlnMetLeu-----LeuArgTrpC 1745  
 Db 5469 CTAAGAAACCTTGAACCTCTCATTCAGGCTGCTCACTTTTGAAGTGAAGAAAGAAA 5528  
 QY 1745 ysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisC 1765  
 Db 5529 CAGATGATGATGAGAA-----GCCATTGTTCTAT-GTGCATGCTTTAACTACT 5578  
 QY 1765 ysPro-----LysLysA 1769  
 Db 5579 GCCCAGATTGTGAAGTGTGTAATTTGTAATCTCCAGTTAATGAGTTTGAAGAAAGAGTC 5638

QY 1769 rgLysIleIleHisAspTyrMetThrAsn 1778  
 Db 5639 TCTGTGCTGCTTCTTCTGTTACTATACAGAT 5667  
 RESULT 2  
 US-08-056-200-93  
 ; Sequence 93, Application US/08056200  
 ; Patent No. 5616500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSES: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (714) 760-0404  
 ; TELEFAX: (714) 760-9502  
 ; INFORMATION FOR SEQ ID NO: 93:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9551 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1507..1644  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 1645..2511  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 2512..8070  
 ; US-08-056-200-93  
 Alignment Scores:  
 Pred. No.: 3.85e-17 Length: 9551  
 Score: 294.00 Matches: 323  
 Percent Similarity: 36.08% Conservative: 291  
 Best Local Similarity: 18.98% Mismatches: 637  
 Query Match: 3.09% Indels: 453  
 DB: 1 Gaps: 66  
 US-09-914-698-1 (1-1861) x US-08-056-200-93 (1-9551)

QY 291 LeuArgLysTyrSerGluSerMetLysAspLeuSerLeuSerProGlnThrLysTyr 310  
 Db 3949 CTAAGCGGAGGACAGAGGAGCGCGAGGAGCGGCTGAAGCGCCAGAGGAGGAGAGAG 4008



5652 GGAGGCGCAATATCCGAAGGAAGAGGAGCTCCACGAGGAGAAGCAGCAGCTGTGTGAGAGA 5711

QY uAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAl 1010

5712 GAAACCGGAGAGAAGAGGCCAGGAGCGGAGAGGCAATGTCTGGGAGGAGCAGGAGCT 5771

QY aAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAspTyrIl 1030

5772 GCAGCAGAGGAGAAGCAGCAGCTGTCTGAGAGAGAACGGGAGAAGAGAGGCCAGGAGCT 5831

QY eGlnMetGlnTrpArg---ThrTyrginLeuGlyArgIleGlnArgHisGluPheLeuAr 1049

5832 GGAGGCGCAATATCCGGAGGAGGAGAGCTTCAGCGGCAGAAAAGGAAACAGCGATACC 5891

QY gGlnArgAspLeuLeuMetPheValGlnArgMetArgSerLysTrpSerMetLeuGl 1069

5892 GGATGAGGAT-----CAGCGCAGTGATCTCAAATTGGCAGTGGGAACC 5933

QY uGlnArg-----LysGI 1073

5934 AGAAAAAGAAAAATCGACTTCGTGATAACAAGTTTACTGCAAGGCCAGAGAGATGAACA 5993

QY uPheGlnGlnLeuLysArgAlaIleAenlleGlnGln-----1086

5994 GTTCCGCGAGTTGGAAGATTCCCAGGTGCGCGACAGACAATCCAGCAAGATCTGCAGCA 6053

QY -----ArgTrpArg-----1089

6054 CTGCTGGGTGAACAGCAAGAGAGAGATCTGAGCAAGAGAGGCGCTGGCAGCAGGC 6113

QY -----AlaLysLe 1092

6114 CAACAGGCAATTCCAGAGGAAGAAACAGCTGGAGCGAGAGAGCAAAAGGAGCCAAG 6173

QY uSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGI 1112

6174 GCGCGACAGAGATGCCAAGAGAAAAGCATGTGCTGAGAGAGAAAGAGAAAGAGAG 6233

QY nAlaTyArgLysAlaThrIleGlnMetArgIleAspArgAsnHistryTy-SerLeuAr 1132

6234 ACCCGCTCAAGACAGACACAGAAAATTCGCGAGGAGGAACAGCTGCTCCAGAAAGGA 6293

QY gLysAsnValIleCyLeuGlnGlnArg-----1141

6294 GGAACAGCGCTCTCCGCCACAGAGCGTCACAGAAAATTCGCGAAGAGAAATTCGCTCCA 6353

QY -----LeuArgAlaIleMetLysMetArgGlnArgGlu---1153

6354 TCAGGAACACAGGAGAGAAAATTCCTCAGAGAGGAACAGCGCGCTGCGGAGAAACGGAGAG 6413

QY -AsnTyLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyArgMetArgGI 1173

6414 AAAATTCCTTAAGAGGAACAG-----CAGCTGCGCTCTCAGAGAGCGCA 6458

QY nGlnMetIleGlnAspArgAsnAlaTyArgTrpArgLysCysIleIleAsnValGI 1193

6459 GCAACTGCGTCAAGACCGGACAGAAAATTCGCGGAGGAGGAACAGCAGCTGAGCGGCA 6518

QY nArgArgTrpArg-----AlaThrLeuGlnMetArgArgGluAr 1206

6519 AGAGCGTGACAGAAAATTCGTCAGAGGAACAGCAGGTGCGCGCCAGCAACAGAGAG 6578

QY gLysAsnTyLeuHisLeuGlnThrThrThrLysArgIleGlnIleLysPheArgAlaLy 1226

6579 AAAATTCCTGGAGAGGAACAGCAGCTGCGCAGGAGCGTCACAGAAAATTCGCGAAGA 6638

QY sArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVa 1246

6639 GGAACAGCTGCTCCAGGAAGGAGAA---CAGCAGCTGCACCGC-----6681

QY lGlnLysArgArgAlaLeuLeuGlnMet-----1256

Db	6682	-CAAGAGCGTGACAGAAAATTCCTGGAGGAGGAACAACAGCTCGCGCCGCAAGAGCGTGA	6748
QY	1257	-ArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrIleLysLeu-----	1273
Db	6741	CAGAAAATTCGCGAACACAGGACTGCCGAGTCAGGAACCCAGAGAGAAAATTCCTCGAGGA	6800
QY	1274	-GlnArgArgPheHisalaGlnLysSerMetArg-----PheMe	1286
Db	6801	GGAAACAGCAGCTGCACCGCCACAGCAACGGCAGAGAAAATTCCTCCAGGAGGAACAGCAGCT	6860
QY	1286	tArgAlaLysTyrArgGlyThrGln-----AlaLaVa	1297
Db	6861	CGCGCGCCAGGAGCGCGGCAACACAGCGCGCTCAGGACCGTCAGAGAAAATTCGCGAGGA	6920
QY	1297	LserCysLeuGlnMethIstTpArgAsnHisLeuLeuArgLysArgGlnArgAsnSerPh	1317
Db	6921	GGACACAGCTGCCCGCAGAGAGGGAGGAACACAGCAGCTGAGCGCCACAGAGCGTCACAGAAA	6980
QY	1317	eLeuGlnLeuArgGlnAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMe	1337
Db	6981	ATTCCGTTTAGAGGAACAGAAAGTGGCGCCCGCAGGAACAAGAGAGAAAATTCATGGAGGA	7040
QY	1337	tIleLysGlnLeuLys-----SerTyrAlaGlnLeuLysGln-----	1349
Db	7041	CGAAACAGCAGCTGCGCCCGCAGGAGGGCCCAACACAGCTCGCGCAGGAGGACAGAAAAT	7100
QY	1350	-----AlaAlaIleThrIleGlnThrArgTyrAr	1359
Db	7101	CCGCGAAGACGAACAGCTGCTCCAGGAAGGGGAAGAACACAGCAGCTGCACCGCCCAAGAGCG	7160
QY	1359	gAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleI	1379
Db	7161	TGACAGAAAATTCCTCGAGGAGGAACCGCAGCTGCGCGCCGAGGAGCGCGAACAA-----	7215
QY	1379	eLysValGlnArgArgTyrArgLysAsnLeuGluMetArgLysGlnIleGluValTyrGl	1399
Db	7216	-----CAGCTCGCTCACGACCGCACAGAAAATTCGCTGAAGAGGAACAGCTGTCTCCA	7268
QY	1399	nLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetAr	1419
Db	7269	GGAAAGGGAGGAACAGCAGCTGCGCGCCGCAAGAGCGGTGACAGAAAATTCGCGGAAGAGGA	7328
QY	1419	gLeyCysLysAlaGlyTyrArgArgIleArg-----LeuSe	1431
Db	7329	ACAGCAGCTCGCGCTCAGGAACGACAGAGAAAATTCCTCCAGGAGGAACACAGCAGCTCGG	7388
QY	1431	rSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluI	1451
Db	7389	CCGCGAAGAACTCGAGAGAAAATTCGCTGAGGAGGAACAGCTGCGCCCAAGAACGGAG--	7446
QY	1451	ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLe	1471
Db	7447	-----CAAGACAGCTCGCGCCGCAAGAACGCTACAGAGAAAATTCCTAGAGGAAGACAGCT	7502
QY	1471	uMetArgGlnInArgArgGluPheGlnMetLysArgArgAlaAlaValIleGlnAr	1491
Db	7503	C-----CGTCCGGAAGGGAAGAACACAGCAGCTGCGCGCCGAGGAGCGC-----GACAG	7550
QY	1491	gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeu---IleGl	1510
Db	7551	AAAAATTCGCGAGGAGGAACAGCTCCGCGCAGGAAGGGAGGAACACAGCAGCTCGCGAGCCA	7610
QY	1510	nSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaAr	1530
Db	7611	AGAGTCT-----GACAGAGAAAATTCGCGGAGGAGGAACACAGCTACGCCGAGGAGAG	7658
QY	1530	gGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnLysLysPheArgGlyLy	1550
Db	7659	GGAGAA---CAGCAGCTCGCCCCCAACACAGCGCTCAGCGAAAGTATCCTGGGAAGAAGA	7715
QY	1550	sArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPh	1570
Db	7716	GCAGCTCCCAACTTGAGAACACAGAG-----CAGAGGCTGGCG-----	7752



Qy	407	Proser-----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProPro	423
Db	4345	CAGGAGGAGAGCGCCACGAGCAACTG---AGGCGCAGCAGCAGGAAAGCGCGCAGCAG	4401
Qy	424	LysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSer	443
Db	4402	CGGCTGAAGCGCGAGGAGGAGGAGAGAG-CCTCGAGCAGCGGCTGAAGCGCGAGCATGA	4460
Qy	444	AlaArgSerSerSerAlaHisAla-----	451
Db	4461	GGAGAGAGGCGCGAGCAGCAGAGCTAGCTAGGAGGAGCAGGAACAGGCCCGGAGCGGAT	4520
Qy	452	--TrpProHis-----AlaGlnSerLysLysPheLysLeuAlaGlnThr	465
Db	4521	TAAGAGCCGCATCCCAAGTGGCAGTGGCAGCTAGAAGCGGAGCGCCAGCAGCAGCGCAAG	4580
Qy	466	MetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrSer-IleGlnPr	485
Db	4581	CAAAAGTCTTACTCGAGGCGCCGCAAGC---AGGAAGGCGAGAGCGCGCAAGAGCAGGA	4637
Qy	485	oSerValLysLeuTyrAspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPh	505
Db	4638	GGAAAAGGCGGCGGAGAGTGAAGTGCATGTCAG-----	4674
Qy	505	eAlaAlaThrThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGluGlnAl	525
Db	4675	-----GAGGAGGAGACG 4685	
Qy	525	aValaSpArgHisGlnAlaAspPheLysLysTyrPheAsnAlaLeuValSerIleProAl	545
Db	4686	GGCTCACCGCAGCAGCAGGAGGAGGAGCGCGCGGATTCACATGGCAGTGGCAGGC	4745
Qy	545	aAspLeuAspAlaAspLeu---AsnAsnLysIleAspValGlyLysLeuPheAsnGluVa	564
Db	4746	GGAGAAAGACGAGGAGGGCGCTCAGAGGCTTCGCCCAGGCCCCCATTCGGAGCA	4805
Qy	564	lArgAsnLysGluLeuValValAlaProThrLysGluGluGlnSerMetAsnTyrLeu--	583
Db	4806	CGGGAGGAGGAGCTG---AGGGCGCAGGAGCGCCAGCAGCGGGAACAACCGTTTCTCC	4862
Qy	584	-----ThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu--	598
Db	4863	GGAGGAGGAGGAGGAGGAGCGCGCGCCGCGCAGCGCGAGGAGGAGGAGGAGGATGCA	4922
Qy	599	-PhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGl	618
Db	4923	GTTCCTCGAGGAAGAGCAGCAGCTCCAGCGCGGGAGCGTGCCCAACAGCTCCAGAGGA	4982
Qy	618	nAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIl	638
Db	4983	GGAGGACGGGCTCCAGAGGATCAG-----	5007
Qy	638	eLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPh	658
Db	5007	-----	5007
Qy	658	eGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLe	678
Db	5008	-----GAGAGGAGGCGCAGGAGCAGCGCCGCGAC-----	5037
Qy	678	uAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuTh	698
Db	5038	-----CAAAATGGAGTGGCACTAGAGAAGAAGAAGAGAGACGCGCCACACGCTG--	5091
Qy	698	rGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuPr	718
Db	5092	-----TAQCCCAAGCCAGCCCTACAAAGCAGCAGCTGAGGAGGAGCAACAGCAGCTGCTG--	5142
Qy	718	oPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheVally	738
Db	5143	-----CAGGAGGAGGAGGAGGAGCTACAGAGAGAGAG-----CG 5177	
Qy	738	sLysSerProHisLysGluThrLysAspIleLeuLeuLeuArgPheSerSerGluLeuLeuAl	758

Db	5178	CGAGAAGAGAAGCGGCCAAGAACACAGGAGAGACACAATACCGCGAGGAACAGCAGCTGCAGCA	5237
Qy	778	aAsnIleGlyAspIleThrArgGlnLeuArgArgLeuGlyTyrValLeuGlnHisArgG	778
Db	5238	CGAGGAGAGACAGCTGCTGAGAGAGAACGG	5279
Qy	778	nThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspG	798
Db	5280	ACGCCAGAGCGGGAAGGCAATAT	5312
Qy	798	yValArgLeuThrArgValValGluValIleLeuLeuArgAspAspLeuThrArgGlnLe	818
Db	5313	TAAGAAGCTGCAGCAGAGAAGACAGCAGCTGCTGGAGAGAACCGGAGAGAAGAGAG--	5370
Qy	818	uArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAl	838
Db	5371	-----CGCCAGAGCGG-----	5382
Qy	838	aLeuGlyGlnAlaAsnPheGlnLeuGlyGlyAspIleAlaAlaGlnAspIleValAspG	858
Db	5383	-----GAGAAAAAATACCGCAGGAGAGAGGATGTCACGACGAGGAA-----	5424
Qy	858	yHisArgGlnLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLy	878
Db	5425	-----GAGCAGCTGCTCAGAGAGAACGG-----	5448
Qy	878	sPheHisAlaAlaAlaThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValVa	898
Db	5449	-----GAGAAAGAGAGCGCCAGAGCTGG-----	5472
Qy	898	lIleGlnArgArgIleAargHisLysGlnLeuMetArgArgHisArgAlaAlaThrValIl	918
Db	5473	-----GAGAGGAGTACCGCAAAAAGACAGCAGCTGCACGACGAGAAGAGAG-----	5517
Qy	918	eGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluAr	938
Db	5518	-----CAGCTGCTGAGAGAGAACG-----	5537
Qy	938	gThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLe	958
Db	5538	GGAGAAAGAGA-----CTCCAGAGCGGGAGAGGCAATATCGGAGAGAGAGAGCT	5591
Qy	958	uTyrGlnSerTyrHisSerIleIleThrIleGlnArg--TrpTrpArgAlaGlnGlnLe	977
Db	5592	GCACGAGGAGAGACAGCAGCTGCTGGAGAGAGAACGGAGACGAGAGGCGCCAGAGCT	5651
Qy	977	uGlyArgGlnHisArg-----GlnArgPheValGlu-----LeuArgG	990
Db	5652	GGAGAGCAATATCGGAAGAGAGAGAGGAGCTGCACGACGAGGAGGACAGCTGCTGAGAGA	5711
Qy	990	uAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAl	1010
Db	5712	GGAAACGGAGAGAGAGCGCCAGAGCGGGAGAGCGCAATGTCGGAGAGAGAGAGCT	5771
Qy	1010	aAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIl	1030
Db	5772	GCACGAGCAAGAGACAGCTGCTGAGAGAGGAACGGGAGAGAGAGAGCGCCAGAGCT	5831
Qy	1030	eGlnMetGlnTrpArg---ThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuAr	1049
Db	5832	GGAGAGCAATATCGGGAGGAGAGAGGTTACGCGCCAGAAAAGGAACACGCAATACG	5891
Qy	1049	gGlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuG	1069
Db	5892	GGATGAGGAT-----CAGCGCAGTGATCTGMAATGCGACGTGGGAACC	5933
Qy	1069	uGlnArg-----LysG	1073
Db	5934	AGAAAAGAAAAATGCAAGTTCGTGATACCAAGGTTTACTGCAAGGCGCAGAGAGATGAACA	5993
Qy	1073	uPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGln-----	1086



Db 5994 GTTCCGCGAGTTGGAAGATTCCAGAGTGGCGGACAGACAATCCAGCAGCAATCTGCAGCA 6053  
QY 1087 -----ArgTrpArg----- 1089  
Db 6054 CTTGCTGGTGAACAGCAAGAGAGAGATCGTGAGCAAGAGAGAGGCGCTGCGCAGCAGGC 6113  
QY 1090 -----AlaLysLe 1092  
Db 6114 CAACAGGCATTTCCAGAGAGAACACAGCTGGAGCGGAGAGCAAGCAAGAGGCAAGCAAG 6173  
QY 1092 uSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGI 1112  
Db 6174 GCGGCACAGGAAGTCCCAAGAGGAAAACAGCTGCTGAGAGAGGAGAAAGAGAGAGAG 6233  
QY 1112 nAlaTyrArgLysAlaThrIleGlnMetMetArgLysAspArgAsnHisTyrTyrSerLeuAr 1132  
Db 6234 ACGCCGTCAAGACAGACAGACAGAAAATTCGCGAGGAGGAACAGCTGCTCCAGGAAAGGGA 6293  
QY 1132 gLysAsnValIleCysLeuGlnArg----- 1141  
Db 6294 GGAACAGCGCTGCTCCGCAAGAGCGTGACAGAAAATTCGCGAAGAGGAACCTGCTCCA 6353  
QY 1142 -----LeuArgAlaIleMetLysMetArgGlnArgGlu-- 1153  
Db 6354 TCAGGAACAGGGAGAAAATTCCTCGAGAGGAACAGCGGCTGCGCAGGAAACGGGAG 6413  
QY 1154 -AsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGI 1173  
Db 6414 AAAATTCCTTAAGAGGAAACAG-----CAGCTGCGCTCGAGGAGCGCGA 6458  
QY 1173 nGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGI 1193  
Db 6459 GCAACTGGCTCAGACCGCGACAGAAAATTCGCGAGGAGGAACAGCAGCTGAGCCGCCA 6518  
QY 1193 nArgArgTrpArg-----AlaThrLeuGlnMetArgArgGluAr 1206  
Db 6519 AGAGCGTGACAGAAAATTCCTGTAAGAGGAACAGCAGCGTGGCGCGCCAGGAAACGAGAGAG 6578  
QY 1206 gLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPheArgAlaLys 1226  
Db 6579 AAAATTCCTGAGGAGGAACAGCAGCTGCGCAGGAGCGTCACAGAAAATTCGCGGAAGA 6638  
QY 1226 sArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuValVa 1246  
Db 6639 GGAACAGCTGCTCAGGAAGAGGAAGAA--CAGCAGCTGCACCGC----- 6681  
QY 1246 lGlnLysArgArgAlaLeuLeuGlnMet----- 1256  
Db 6682 -CAAGAGCGTGACAGAAAATTCCTGGAGGAGGAACAAACAGCTGGCGCCCAAGAGCGTGA 6740  
QY 1257 -ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeu----- 1273  
Db 6741 CAGAAAATTCGCGCAACAGGAACCTGCGCAGTCCAGGAACCCAGAGAAAATTCCTCGAGGA 6800  
QY 1274 -GlnArgArgPheHisAlaGlnLysSerMetArg-----PheMe 1286  
Db 6801 GGAACAGCAGCTGCACCGCCAGCAACGCGCAGAGAAAATTCCTCCAGGAGGAACAGCAGCT 6860  
QY 1286 tArgAlaLysTyrArgGlyThrGln-----AlaAlaVa 1297  
Db 6861 GCGCGCCGAGGAGCGCGGCAACAGCGCGTCCAGGAGGAACAGAAAATTCGCGGAGGA 6920  
QY 1297 lSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPh 1317  
Db 6921 GGAACAGCTGGCGCAGGAGAGGAGGAACAGCAGCTGAGCGCGCCCAAGAGCGTGACAGAAA 6980  
QY 1317 eLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMe 1337  
Db 6981 ATTCCGTTAGAGNACAGAAAGTGGCGCGCCAGGAACAGAGAGAAAATTCATGAGGA 7040  
QY 1337 tIleLysGlnLeuLys-----SerTyrAlaGlnLeuLysGln----- 1349  
Db 7041 CGAACAGCAGCTGCGCGCCAGGAGGGGCCAACAAACAGCTGGCGCCAGGAGGAGCAGAAAAT 7100

QY 1350 -----AlaAlaIleThrIleGlnThrArgTyrAr 1359  
Db 7101 CCGCGAAGACGAACAGCTGTCTCCAGGAAGAAAGAACAGCAGCTGCACCGCCAGAGCG 7160  
QY 1359 gAlaLysLysAlaMetGlnLysGlnValLeuTyrGlnLysGlnArgGluAlaIleI 1379  
Db 7161 TGACAGAAAATTCCTCGAGGAGGAACCGCAGCTGCGCGCCAGGAGCGGAACAA----- 7215  
QY 1379 eLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleValTyrGI 1399  
Db 7216 -----CAGCTGGCTCAGCAGCGCGACAGAAAATTCGTCGAGAGGAACAGCTGCTCCA 7268  
QY 1399 nLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetAr 1419  
Db 7269 GGAAGGGAGGAACAGCAGCTGCGCGCCCAAGAGCGTGACAGAAAATTCGCGGAGAGGA 7328  
QY 1419 gLeuCysLysAlaGlyTyrArgArgIleArg-----LeuSe 1431  
Db 7329 ACAGCAGCTCGCGCTCAGGAACGAGAGAAAATTCCTCCAGGAGGAACAGCAGCTGC 7388  
QY 1431 rSerLeuSerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGlnArgGluI 1451  
Db 7389 CCGCCAGGAACCTGAGAGAAAATTCCTGTGAGAGGAACAGCTGCGCCCAAGAAAACGGAG-- 7446  
QY 1451 ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLe 1471  
Db 7447 ----CAAGACAGCTGCGCGCCCAAGAACGCTACAGAAAATTCCTAGAGGAGAGCAGCT 7502  
QY 1471 uMetArgGlnGlnArgArgGluPheGluMetLysArgArgAlaAlaValIleGlnAr 1491  
Db 7503 C---CGTCCGGAAGGAGGAACAGCAGCTGCGCGCCAGGAGCGC-----GACAG 7550  
QY 1491 gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeu---IleGI 1510  
Db 7551 AAAATTCGCGAGGAGGAACAGCTCCCGCAGGGAAGGAGGAACAGCAGCTGCGCAGCCA 7610  
QY 1510 nSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaAr 1530  
Db 7611 AGAGTCT-----GACAGAAAATTCGCGAGGAGGAACAGCTTACCCAGGAGAG 7658  
QY 1530 gGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnLysPheArgGlyLy 1550  
Db 7659 GGAAGAA---CAGCAGCTGCGCCCAACACAGCTGCACGGAAGTATCGCTGGGAAGA 7715  
QY 1550 sArgLeuMetIleGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPh 1570  
Db 7716 GCAGTCCAACTTGAGGAACAAAGAG-----CAGAGGCTGCGG----- 7752  
QY 1570 eGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGI 1590  
Db 7753 -CAGAGCGAGACCGCAGTACCGCGGCGGAGAGCAGTTGCCACG----- 7797  
QY 1590 uMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGI 1610  
Db 7798 -----CAGGAGAGAGTCTGTCGTGAGGAACAAGAACTATGGCAAGA 7838  
QY 1610 yTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAl 1630  
Db 7839 AGAGGAGCAGAAAACGTCGCCAGGAACGGGAAGAAAATTCAGGGAAGAAACATCCGCGC 7898  
QY 1630 aGlnLeuArgGlnGlnAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaVa 1650  
Db 7899 CCAGACAGAGGAACAGAGGCAACCCCAAGTCGGGAGGATACATCCCAAGAAAGGAA 7958  
QY 1650 l-----ArgPheLeuArg-----GlyArgPheIleAla-SerAspAlaLeuA 1664  
Db 7959 GGGCCATGGGCGGCTTCTGGAGCCCGGCACATCATAGTTTCCAGTGTCCCGAGTGGCGTC 8018  
QY 1664 laValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1684  
Db 8019 CAGCCCTCTCTATGAGTACATCCAAAGACAGAGATCTCAATACCAGCCCTTAAGTGTGTT 8078







2734	Db	GTTCAGAAAGACTAAATGAGATGGAAACAGCTCAAGGAAACAATAGAAAATAGAGATTCT	2793
276	Qy	AlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTyrSer	285
2794	Db	CCGCTGCAAACTGTGAGAGGGGAAACACTGATTACTCAGAAACTGCAGCAAACTTTA	2853
296	Qy	GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySer	315
2854	Db	GAGAACTAAAACTTTAACT	2874
316	Qy	MetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlu	335
2875	Db	CAAGAAAAGAGATGATCTAAACCAACTCCAGAAAGCTTGCAAATGAGAGG	2925
336	Qy	GlnGlnIleGlnIleLysAlaLysAsp	354
2926	Db	GACCACTCAAAAGTGATATTCAGCATCTGTTAAACATGATATAGATACTCAAGAACAA	2985
355	Qy	LeuAlaGlyGlnGlnGluPheLeuPheAsnHis	366
2986	Db	TTACGAAATGCTCTTGAGTCTCTGAAACCAACATCAAGAAACAATTAATACACTAAATCG	3045
367	Qy	GluIleLeuAlaGlnSerSerArgPheAsnLeuHis	378
3046	Db	AAAATTTCTGAGGAAGTTTCAGG	3102
379	Qy	GluValGlyArgLysSerValLysGlySerProValLysAsnProHisLysArg	396
3103	Db	AAAGATGATTTTCAGCAAAAGATGTT	3159
397	Qy	ArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyrArgAsnGlu	416
3160	Db	AAAAATACCCAAACACTAACTGCAGATGTTAAGGATAATGAGATAATGAGCAAAACAAAGG	3219
417	Qy	ThrVal	421
3220	Db	AAGATATTCTTTAATACAGGAGAAAATCAACTCCAAACAATGTTAGAGAGTGTTATA	3279
422	Qy	ProProLysLysGln	430
3280	Db	GCAGAAAAGGACAAATGGAAGCTGACCTAAAGGAAATATGTAATGACCATTTGAAAC	3339
431	Qy	ThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHis	450
3340	Db	CAGGAAGATTAAGACTCTCTGGGATCAACTTAAAAAGCAACAAGAGATATGTCACAA	3399
451	Qy	AlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLys	470
3400	Db	GAAAAAGAACCATGCCATTAAGAAAGAGAGAGCTTCTTAGCACTGTGACAGACTGCCA	3455
471	Qy	LysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLysLeuTyr	490
3460	Db	GAA	3507
491	Qy	AspSerGluLeuTyrMetGlnThrCysLeuAsnProAspProPheAlaAlaThrThrThr	510
3508	Db	CAGCAACAACCTTCTTAAT	3522
511	Qy	IleAspProPheLeuAlaSerThrMetTyrLeuAspGluAlaValAspArgHisGln	530
3526	Db	GTCAAGAAAGAGATG	3544
531	Qy	AlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAsp	550
3541	Db	AGTGATGCAGAAAAGATTAATGAATA	3583
551	Qy	LeuAsnAsnLys	559
3589	Db	TTAAAGAACAAAGAAATGACATTGGAAACATATGGAAACAGAGAGGCTTCAGTTGGCTCAG	3644
560	Qy	LeuPheAsnGluValValArgAsnLysGluLeuValValAlaProThrLysGluGlnSer	579

Db 3649 AAACCTTAATGAA-----AATTATGAGGAAGTGAAATCTATAACCAAGAAAGAAAGTT 3702  
QY 580 MetAsnTyrLeuThrLys  
::: |||  
Db 3703 CTAAGGAATTAACAGAAGTCATTTGAAACAGAGAGAGACCACCTTAGAGGATATATAAGA 3762  
586 -----TyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu 598  
Db 3763 GAAATTGAAGCTACAGGCTCAACACCAAGAAAGAACTAAAAATTCATATTCACCTA 3822  
QY 599 -----PhePheSerGluGlnMetArgLeuProCysSerLysValAlaVal 613  
Db 3823 AAAGAACACCAAGAAACTATTGATGAACCTAAGAGAACGGTATCTGAGAGACAGCTCAA 3882  
QY 614 TyrValAsnLysGlnAlaLeuArgLysArgSerAspArgAsnLeuHisLeuAspValVal 633  
::: |||  
Db 3883 ATATAAATACTCAGGACTTAGAA---AAATCCCATACCAATTTACAAGAGAGATCCCA 3939  
QY 634 MetGlnArgThrLysLeuGluLeuLeuLeuCysPheAsnProLeuTyrLeuArgLeuGly 653  
::: |||  
Db 3940 GTGCTTCATGAGGAACAAGAGTTACTG-----CCTAATGTGAAAAAAGTCAGT 3987  
QY 654 -----LeuGluValValPheGlyGluLysLysLysLysLysLysLysLys 665  
Db 3988 GAGACTCAGGAACAATGAATGAATGAGTTATTA-----ACAGAACAGTCC 4035  
QY 666 SerAsnArgAspLysValGlyLeuSerThrPheLysLeuAsnArgLeuPheArgAsnLys 685  
::: |||  
Db 4036 ACAACCAAGACTCAACACACACTGGCAAGATAGAAATGGAAGGCTCAGTTGAATGAA 4095  
QY 686 -----CysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTyrAlaGlu 703  
Db 4096 AAATTCACAGAAAGTCAGGAAGAGATATAATCTCTACCAAGGAAGAAAGACAACCTTAAA 4155  
QY 704 ThrLysLys-----LysHisSerLeuGlnLysLysLysLysLysLysLysLys 718  
Db 4156 ACGATAAAGAGCCCTTGAGTTAAACAT----- 4185  
QY 719 PheLeuAspGlnAlaLys-----GlnLysArgLys 728  
Db 4186 -----GACAGCTGAAAGAACATATTAGAGAACTTTGCTTAAATCCAGGAGTCTCAA 4239  
QY 729 ValLysHisAsnProCysLeuPheValLysLysSerProHisLysGluThrLysAspLys 748  
::: |||  
Db 4240 AGCAAAACAAGACAGTCTTAAATATGAAAGAAAGAAAGACAATGAACACTACCAAAATCGTG 4299  
QY 749 -----LeuLeuArgPheSerSerGluLeu 756  
Db 4300 AGTGAGATGGAGCAATTCAAACCAAGATTTCAGCTACTAAGGATAGAAATAGAAATG 4359  
QY 757 LeuAla-----AsnLysGlyAspLys 763  
Db 4360 CTGGGATTTGTCAAAAGAGACTTCAAGAAAGTCATGATGAATGAATCTGTAGCTAAGGAG 4419  
QY 764 ThrArgGluLeuArgArgLeuGlyTyrValLeuGlnPheHisArgGlnThrPheLeuAspGlu 783  
::: |||  
Db 4420 AAAGATGACTACAGAGGCTGCAAGAAAGTCTTCAA-----TCTGAA 4461  
QY 784 PheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrArg 803  
Db 4462 AGTGACCACTCAAAGAAACATATAAGAAATTTAGCTTAACACCTGGAAACTGAAGAG 4521  
QY 804 ValValGluVal-----LysLeuLeuArgAspAspLeuThrArgGlnLeuArg 819  
Db 4522 GAACTTAAAGTTTGCTCATTTGTTGGCTTGAAGAAACAAGAGGAACTATTATTCAGTTAAGA 4581  
QY 820 ValProAlaLysSerArgLeuGlnArgLysPheAsnValLysLeuAlaLeuGlyAlaLeu 839  
::: |||  
Db 4582 GTGAATCTTCAGAGAGGAAGTGAATATCAACCATTCAAAGCAGTTAGAACCAATC 4641  
QY 840 GlyGluAlaAsnPheGlnLeuGlyGlyAspLysAlaAlaGlnAspLysValAspGlyHis 859  
::: |||  
Db 4642 -----AATGATAAATTACAGAACAGATC-----CAAGAGATTATGAGAAAGAG 4686

QY 860 ArgGlu-----LysThrLeuSerLeuLeuTyrGlnLeuLeuLysPheArgSer 876  
::: |||  
Db 4687 GAACAACTTAATAATAAAACAAATTAGTGAGGTTTCAGGAAAAAGTGAATGAATGAACAA 4746  
QY 877 ProLysPheHisAlaAlaAla-----ThrValLeu 886  
Db 4747 TTCAGGAGCATCGCAAGCCCAAGGATTCAGCACTCAAAAGTATAGAAAGTAAAGTATGCTC 4806  
QY 887 GlnLysTyrTyrArgHisTyrLeuHisValValLysGlnArgArgLysArgHisLys 906  
::: |||  
Db 4807 GAGTTGACCAACAGACTTCAGAAAAAGTCAAGAGAAATATACAAATATGATTAAAGGAAAA 4866  
QY 907 GlnLeuMetArgArgHisArgAlaAlaThrValLysGlnAlaValPheArgGlyHisGln 926  
::: |||  
Db 4867 GAGGAATGAAAGAGTACAGGAGGCCCTTCAGATAGAGAGAGACCAACTGAAAGAAAAAC 4926  
QY 927 MetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaAlaLysLysLys 946  
::: |||  
Db 4927 ACTAAAGAAATTTAGCTAAATGAAAGTCTCAAGAAAAAGAAATATCAGTTCTTT 4983  
QY 947 LysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerLysLys 966  
::: |||  
Db 4984 AAGATGACAGCTGTCAATGAGACTCAGGAGAAAAATGTGTGAAATAGAACAC 5034  
QY 967 ThrLysGlnArgTyrTyrArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVal 986  
::: |||  
Db 5035 -----TTGAGGAGGAGCAATTTGAGACCCAGAGTTAAACCTGGAACACATAGAAACG 5088  
QY 987 GlnLeuArgGluAlaAlaLysPheLeuGlnArgLysTyrArgArgArgLeuPheAlaLys 1006  
::: |||  
Db 5089 AATATAAGTTGACTCAGATACATACATGAAACCTTGAAGAAATGAGATCTGTAAACAAA 5148  
QY 1007 Lys-----LeuLeuAlaAlaLysGlnThrAlaArgLeuGlnArgSerGln 1021  
::: |||  
Db 5149 GAAAGAGATGACCTTAGGAGTGTGAGAGACTCTCAAAGTAGAGAGAGCAGCTCAAG 5208  
QY 1022 -----LysGlnGlnAlaAlaLysTyrTyr 1030  
Db 5209 GAAACCTTAGAGAACTATACTAGACCTAGAAAAACAGAGGGAGCTTAAAAATGTT 5268  
QY 1031 GlnMetGlnTyrArgThrTyrGln----- 1038  
Db 5269 CACATGCACTGGAAGGAGCACCACCAAACTATTGATAAACTAAGAGGATTTCTTCAGAG 5328  
QY 1039 -----LeuGlyArgLysGlnArg-----HisGluPheLeuArg 1049  
Db 5329 AAAACAAATGAAATATCAATATGCAAAAGGACTTAGAAACACTCAAAATGATGCTTAAAA 5388  
QY 1050 GlnArgAspLysLysMetPheValGlnArgArgMetArg---SerLysTyrSerMetLeu 1068  
::: |||  
Db 5389 GCACAGATCTG-----AAATATCAAGAGGAACTAAGAAATGCTCACATGCACTGAAA 5442  
QY 1069 GluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaLysLysGlnGlnArgTyr 1088  
::: |||  
Db 5443 GAGCAGCAGGAACTATTGACAAACTCAGA-----GGAATTTCTTCTGAGAAG 5490  
QY 1089 ArgAlaLysLeuSer---MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSer 1107  
Db 5491 ACAGATAAACTATCAATATGCAAAA-----GATTTAGAAATTTCA 5532  
QY 1108 ValLeuLysValGlnAlaTyrArgLysAlaThrLysGlnMetArgLysPheArgHis 1127  
Db 5533 AATGCTAAATTAACA-----GAAAGATTCAGAACTTAAGGCAATGAAATCA 5583  
QY 1128 TyrTyrSerLeuArgLysAsnValLysCysLeuGlnGlnArgLeuArgAlaLysMetLys 1147  
::: |||  
Db 5584 CTTATTACCTTAAAAAAGATGCTCAATGAGACACAGAAAAAGTGTCTGAAATGGAGCAA 5643  
QY 1148 MetArgGluGln---ArgGluAsnTyrLeuArgLeuArgAsnAlaSerLys 1163  
::: |||  
Db 5644 CTAAGAAACAAATAAAGAACCAAGCTTAACCTGAGTAAATTAGAAATAGAGATTTA 5703

QY 1164 ---LeuValGlnLysArgGlyArgMetArgGlnGlnMet-----IleGlnAspArg 1179  
Db 5704 AATTGGCTCAAGAACTTCATGAAGAACTTCGTAATGAAAGAAAGA 5763  
QY 1180 AsnAlaTyrLeuArgThrArgLysCysIle----- 1189  
Db 5764 GATAATCTAAGAGAGTAGAGACACTCAAACTGGAGAGAGACCAACTCAAGGAAGC 5823  
QY 1190 -----IleAsnValGlnArgArgGlyArgGlnAlaThrLeuGln 1201  
Db 5824 CTGCAAGAAACCAAGCTAGAGATCTGAAATCAACAGGAACCTAAAGCTGCTATG 5883  
QY 1202 MetArgGlnArgLys-----AsnTyrLeuHisLeuGlnThrThrLysArg 1218  
Db 5884 CTATCAAGAAACCAAGAACTGTTGATAAAGCTTTAGAGAAATTTTCAGAAAGACA 5943  
QY 1219 IleGlnIleLysPheArgAlaLys----- 1226  
Db 5944 ATTCAATTTTCAGACATTCAGAAAGGATTAGATAAATCAAAAGATGAATTACAGAAAAAG 6003  
QY 1227 ---ArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThrLeuVal 1245  
Db 6004 ATCCAAGAACTTCAGAAAGAACTTCAACTGCTTAGAGTCAAAAGAGATGTCATATG 6063  
QY 1246 ValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGln----- 1261  
Db 6064 AGTCATAAAAAATTAATGAATGGAACAGTTGAAGAAGCAATTTGAGCCAAAACCTATCTA 6123  
QY 1262 -----GluTyrLeuHis-----LeuArgGlu 1268  
Db 6124 TGCAAGCTGAGATGATAAATTCAGTTGACTAAGAACTTCATGAAGACCTTGAAGAA 6183  
QY 1269 ValThrIleLysLeuGlnArgPheHisAlaGlnLys-----SerMetArgPhe 1285  
Db 6184 ATAAGAATTGTAGCTAAAGAAAGAGATGAGCTAAGAGGATAAAAGAACTCTCAAAATG 6243  
QY 1286 MetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeuGlnMetHisTrpArg 1305  
Db 6244 GAAAGGACCAATTCATAGCAACCTTAAGGAAATG-----ATAGCTAGAGACCGACAG 6297  
QY 1306 AsnHisLeuLeuArgLysArgGlnArgAsnSerPheLeuGlnLeuArgGlnAlaLys 1325  
Db 6298 AACCACCAAGTAAACCTGAAAGAGG-----TTACTAAGTGTGACACACAG 6345  
QY 1326 ThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGln---LeuLysSerTyr 1344  
Db 6346 CACCTTATGGAAGCCTCAGAGAAAGTCTCTAGATAAAGAGCTTTTGAAGAGATAC 6405  
QY 1345 AlaGln-----LeuLysGlnAlaAlaIleThrIleGlnThrArg 1357  
Db 6406 TCAGAGATGGATGATCATATGATGCTTGAATAGATTGCTCTGACTTGGAGAGGNA 6465  
QY 1358 TyrArgAlaLysAlaMetGln-----LysGlnValValLeuTyr----- 1371  
Db 6466 ATTGAATTCACAGAACTCATGAAGAACTGAAGTATGTTTGAAGCTATGTTACAAAAATA 6525  
QY 1372 ---GlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGlyAsnLeuGlu 1390  
Db 6526 AAAAGAGAACCAACATCATCAATAAATTTGAATGGATTT----- 6570  
QY 1391 MetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLys 1410  
Db 6571 ---ATTGATGAAGTGGAAAGCAAGAAATTCGTAATTAATAATACAGCAC 6618  
QY 1411 TrpTrp-----ArgSerIleArgAspMetArgLysCysAla 1423  
Db 6619 CTTCAACAGATGTGATGATCATCCAGAGATTAAGGAACTCAAAATGAACACAG--- 6675  
QY 1424 GlyTyrArgArgIleArgLeuSerSerLeu----- 1433  
Db 6676 ---AATATGATCTACATATGAGGAAATTTCTCAAGATTTCTCAGAAAGATGATGCCCT 6732  
QY 1434 SerIleGlnArgLysTrpArgAlaThrValGlnAlaArgGlnArgGlnIlePheLeu 1453

Db 6733 AGCATAAGACTGAATTTCAACAAGTACTAAGTAATAGGAAAGAAATGACACAGTTTGTG 6792  
QY 1454 Ser-----ThrIleArgLysVal----- 1459  
Db 6793 GAAGAGTGGTTAAATACTCGTTTTGTATATAGAAAGCTTAAATAATGGCATCCAGAAAGAA 6852  
QY 1460 -----ArgLeuMetGlnAla-----PheIleArgAlaThrLeuLeuMetArgGln 1474  
Db 6853 AATGATAGATTTCAGTCAATAAATCTTTTAAACAGAAATAATGGCATAATGAAT 6912  
QY 1475 GlnArgArgGluPheGluMetLysArgAlaAlaValIleGlnArgPheArg 1494  
Db 6913 GAATCAACAGAGTTTGAGGAAAGA-----AGTGTACCATATCCAAAGAGTGGGAA 6963  
QY 1495 AlaArgCysAlaMetLeuLysAlaArg-----GlnAspTyrGlnLeuIle 1509  
Db 6964 CAGGACCTGAAATCACTGAAAGAGAAAAATGAAAACTATTTAAAACTACCAACATTTG 7023  
QY 1510 GlnSerSerValIle-LeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAl 1529  
Db 7024 AAGACTTCCTTGGCATCTGGTCCAGGTTAATCTCTACCAACAAGACA----- 7072  
QY 1529 aArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu----- 1543  
Db 7073 -----ATAAGAAATCTCTCATGTATCATCAAGAGCTACACAGTTAAACC 7113  
QY 1544 -GlnGlnLysPheArgGlyLysArgLeuMetIleGlnArgAsnCysPheGlnLeu 1563  
Db 7114 ACAGAGAAATTCGAGAGCTGGAATAAT-----CACTCGTAGAAGCTAAA 7158  
QY 1563 uArg-----CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLys 1581  
Db 7159 GAAAGCTGTATGCATA-----AGGAACGACAGATTATAAAGATGCAGAAA 7203  
QY 1581 sArgPheGlnAlaLeuMetThr-----ProGluMetMetAspLeuI 1595  
Db 7204 GAACCTTGAGTGACTTAATGACATAATGCAAAAATTCAGGCCAAAGTTTCATGAATCAAT 7263  
QY 1595 eArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIle-Arg- 1614  
Db 7264 AAATGCCCTTGAAGAAACAAAGAGACAATCAAGTACTTCAGGACAAAGTTGCTTTAGGA 7323  
QY 1615 ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeu----- 1632  
Db 7324 GCTAAGCCATATAAGAGAAATTCGAAGATCTCAAAATGAAGCTTGTGAATAGACCTA 7383  
QY 1633 -----ArgGlnGluAlaLysAlaValAsnSerValArgCysLysVal 1646  
Db 7384 GAGAAATGAAAAATGCCAAAGAAATTTGAAAGGAAATCAGTGTACAAAAGCCACTGTA 7443  
QY 1647 Gln-----GluAlaValArgPheLeuArgGlyArgPhe-----IleAla 1659  
Db 7444 GAATATCAAAAGGAGATTATAGGCTATTGAGAGAAATCTCAGAAAGATCAACAGGCC 7503  
QY 1660 SerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeu 1679  
Db 7504 CAAATACCTCAGTGTATCAGACATACTGATCTCAGCCTTCAATAAACC----- 7557  
QY 1680 LeuMetTrpCysSerGlu-----PheMetSerThrPheCysTyrGlyIleMet 1695  
Db 7558 ---TTAACTTGTGGAGGTGGCAGGCAATTTGACAAAACACAAAGCTCTTTATTTGAAA 7614  
QY 1696 AlaGlnAlaIleArgSerGluValAsp-----LysGlnLeu 1707  
Db 7615 AGTGACATATAAGGCTAGAAAAAGAAATTTCTAAGTTAAAGCAGCAAAAATGAACAGCTA 7674  
QY 1708 IleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsn 1727  
Db 7675 ATAAACAAAGAAATGAATTTGTAGCAAT---AATCAGCATCTTCCATGAGTCAAA 7731  
QY 1728 ThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeuArgTrpCysAspLys 1747



```

Db      7732 ACTTGGAGGAA-----AGAACCCCTTAAAGAGAGGCTCACAAA 7770
QY      1748 AspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPheAlaHisCysProLys 1767
Db      7771 CAAGTAACCTGTGAGATTCT-----CCAAG 7797
QY      1768 LysArgLysIleleHisAspTyrMetThrAsnProGluAlaIleTyrMetValArgGlu 1787
Db      7798 TCTCCTAAAGTGAAGTGA-----ACAGCTTCT 7824
QY      1788 ThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsnAlaArgLysProPro 1807
Db      7825 AAAAGAAACAATACACCTCTCAATGCAAGACGGAATTTACAAAGATCCTGTGCCA 7884
QY      1808 MetThrSerGlyArgTyrLysSerGlnLysIleAsnPheThrProCysSerLeuProSer 1827
Db      7885 AAGGAATCACCA-----AAATCTGTGTTTTGATAGCCGATCAAGATCTTTACCAATCA 7938
QY      1828 LeuGluProAspPheGlyIleleArgTyrSerProTyrThrPhe-IleSerSerValty 1847
Db      7939 CCTCATCCA-----GTTGCTAT-----TTTGATAACTCAAGTTTA 7974
QY      1847 rAlaPheAspThrIleLeuCysLysLeuGln 1857
Db      7975 GGCTTT---GTCAGAGGTGCAAAATGCAG 8002

RESULT 6
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitulum Genome, Fragment
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Alignment Scores:
Pred. No.: 5.93e-12 Length: 580073
Score: 273.50 Matches: 304
Percent Similarity: 36.58% Conservative: 271
Best Local Similarity: 19.34% Mismatches: 598
Query Match: 2.87% Indels: 400
DB: 4 Gaps: 71

US-09-914-698-1 (1-1861) x US-08-545-528D-1 (1-580073)
QY      296 GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyr----- 310
Db      259797 GAGAAATCAAAAGAGATTATTAAGTATTGAGTATGAAATACCTATATAGACAGTAGTAGT 259856
QY      311 -----AlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerIle 327
Db      259857 GCAGATAATCACTGCAAAATGTTTATGAAACATCGATCAAAATCAGATCCAGTTTAAA 259916
QY      328 GluGlnAsnArgTyrTyrGln-----GluGlnGlnIleGlnIleLys 341
Db      259917 CACCAATACCAAACTTATAGAGATGAGTTAAGTCAACTTGGCGCAAGATCCAGCTCACC 259976
QY      342 AlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGlnGluPhe 361

```

```

Db      259977 AAACAAGAGTTAGTTGATAAAGAAATCAGCACTAAGAGTAAGAGTAGATGCTGTAATTT 260036
QY      362 LeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGly 381
Db      260037 TACATTAAATCCCGT-----TTAGCTGAA-----CTTGATGATGTAGCT 260075
QY      382 ArgLysSerValLysGlySerProValLysAsnProHisLysArgArgSerHisGluLeu 401
Db      260076 AAA-----CAACTT 260084
QY      402 SerPheSerAsp-----AlaProSerAsnGluSerLeuTyrArgAsnGlnThrValAla 419
Db      260085 AGCTTTCAAGATGGTATTACCAAGCAAAATGCTCAACATCTTGAGGATAAGTTAGTTGCT 260144
QY      420 IleSerProLys-----LysGluArgValGluAspThrThrLeuProArg 435
Db      260145 TTAACAAGAAAAAGAACCCGTTTAAATACCCAAAAGAGCGCTTTTAACTTAAAGACAA 260204
QY      436 SerAla---AlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHis 454
Db      260205 TCTGCTTTAATTCATATCAATAAATCCACAGAAAAATGAACTGTTGCT---AAGCAC 260261
QY      455 Ala-----GlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLys----- 470
Db      260262 TTAGAAACACCAAGCAAAATGAGTTTGAACAAAAACAGCTCTGATAGCTCTTTAAAGCTGAA 260321
QY      471 -----LysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLys 488
Db      260322 ACTGAATATAAGGCTTTTACAACAACAAGATTAAATGAAATTTAAAAAATGAGTGCACATAAG 260381
QY      489 LeuTyrAspSerGluLeuTyrMetGlnThrCysIleAsnProAspProPheAlaAlaThr 508
Db      260382 -----AGCGAGGAATTTTAAACCAA-----GAACGGGAACATTTTGAACACGC 260426
QY      509 ThrThrIleAspProPheLeu-----AlaSerThrMetTyrLeuAspGlu----- 523
Db      260427 AGGGAATTTGACACGCTTTTAAACCAAGCATCTTTGGAATATGAACACCAACGTGAGTCA 260486
QY      524 ---GlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAlaLeuValSer 542
Db      260487 AGTCAACTACTCAAGATAAGCAGATGAAGTAAACCAACACTTCCAAACCTTAGAGTAT 260546
QY      543 IleProAlaAspLeuAspAlaAspLeuAsn-----AsnLysIleAspVal 557
Db      260547 GCTAAAAGAGGCTAGACAAAGAAAGAACCTTTAGATCAACAAAAAAGTTGATAGT 260606
QY      558 GlyLysLeuPheAsnGluValArgAsnLysGluLeuValValAlaProThrLysGluGlu 577
Db      260607 GAAGCAATCTTT-----CAACTCAAGAAAAAGTTTCTCAAGAACCTAAGGAGCTT 260657
QY      578 GlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGlu 597
Db      260658 GAAGAACTA---TACCTTGTAAAAAACAACCAAGATCAGAAAGGAAAAATGAATG--- 260711
QY      598 LeuPhePheSerGluGlnMetArg-----LeuProCys 608
Db      260712 TTGTTTTTGAAGACGAGTTAAACACACCAACAGCAGATTTGAAATGAATGGAGCT 260771
QY      609 SerLysValAlaValTyrValAsnLysGlnAlaLeuArgIleArgSer----- 624
Db      260772 AAACAACAGGAGTTGTTGAAGCTAAACATGCAATGGAA---CGTTCTTTTATCAAACTT 260828
QY      625 ---AspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluLeuLeu 643
Db      260829 GAAGATAAGAAAAAGATCTTAACCACTAAGCACACACAG----- 260867
QY      644 CysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlyLysIleGln 663
Db      260868 -----ATTGCCAATGAGTTTCCCAA 260888
QY      664 MetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPheArg 693

```



Db 260889 CTAAAAAGTGAAGTCAAGAGCTGCTGATTTTGAACATAATGTTGCCAAATGAGTATGAA 260948  
 Qy 684 Asn---LysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrGluGluTyrAla 702  
 Db 260949 AACTTGCACCAAGAAAAACAAAGATTATTCGAAGAACGTAAGTACTTACTTCTTGAAGGAATGCT 261008  
 Qy 703 GluThrLeuLysLysHisSerLeuGlnLysLeuPheLeuLeuProPheLeuAspGln 722  
 Db 261009 GCGGTTTATCAAAACCGGTTTACAAACAAAACGTCGAGGAGTTATTA----- 261053  
 Qy 723 AlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerProHis 742  
 Db 261054 -----CAACAA 261059  
 Qy 743 LysGluThrLysAspIleLeuLeuArg---PheSerSerGluLeuLeuAlaAsn----- 759  
 Db 261060 AAAGAACGCTTGATCAGCTTACAAAAGCTTTGAGCAAGAACGCTTAAATCAACCAAGG 261119  
 Qy 760 -----IleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeu 774  
 Db 261120 GAACACAAAGAGTGTGGTTCATCAGTTGAAAACAAAAGAGATATTCGGCAAAAAATC 261179  
 Qy 775 Gln---HisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeuAla--- 792  
 Db 261180 CAAGATTTTCCCAACTTCACTGAAC-----GCTTCTAAAAATTTAGCTGAA 261227  
 Qy 793 -----ValAspLeuAspGlyValArgLeuThrArgValValGluValIleLeuLeu 810  
 Db 261228 CGGGAGATGGCAATCAAGTTTAAAGAAAGGAGATAGAGCAACTGAAAAGCAACTGTTA 261287  
 Qy 811 ArgAspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePhe 830  
 Db 261288 AATGAT-----GTTAAT 261299  
 Qy 831 AsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAsp--- 849  
 Db 261300 AATGCTGAAGTTATCCAGCACACTTAGCACACTCAACCAATCACTTAACCAAGACGC 261359  
 Qy 850 -----IleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSer 865  
 Db 261360 AGTGAATTCAAAACGCCAAACAAAGGATGTGCTGATTTTCAACAATGATTCACATAAAAAA 261419  
 Qy 866 LeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrVal 885  
 Db 261420 CTC----- 261422  
 Qy 886 LeuGlnLysTrpTrpArgHisTrpLeuHisValIleGlnArgAlaThrValIleGlnAlaVal 905  
 Db 261423 -----AATGAGTATGAACCTTAGCTTACAAAACGGTTGCAAGAA 261461  
 Qy 906 LysGluLeuMet-----ArgArgHisArgAlaAlaThrValIleGlnAlaVal 921  
 Db 261462 TTACAAACCCCTTGAGGCTTAACCAAAAACCAATTCATATCAAAAT-----CAAGCTTAC 261515  
 Qy 922 PheArgGlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAla 941  
 Db 261516 TTGTAAGGT-----GAACTTGATAAATCTTAACGAGAGAAAACAA 261554  
 Qy 942 AlaIle-----IleLeuGlnLysPheThrArgArgTyrIleAlaGlnLysGlnLeuTyr 959  
 Db 261555 GCTTTTGTGAACCTTAGTAAAGAAACAACTATGAGGTTGATCTATTAACAAAGGTTG 261614  
 Qy 960 GlnSerTyrHisSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArg 979  
 Db 261615 AGTGATAAACATCAAGCTTTAAATATGCAACA-----GCAGAGCTAGATAGA 261662  
 Qy 980 GlnHisArgGlnArgPheValGluLeuArgGluAlaIlePheIleLeuGlnArgIleTrp 999  
 Db 261663 AAAACCCAT-----GAGTTAAATAATGCT-----TTTTAAACCATGATGCG 261704  
 Qy 1000 ArgArgArgLeuPheAlaLysLysLeuAlaAlaAlaGluThrAlaArg----- 1016  
 Db 261705 GATCAAAAGAGTCTACAGGACCACTAGCACTGTTAAAGAGACCCCAAACTAATTGAT 261764

Qy 1017 LeuGlnArgSer-----GlnLysGlnGlnAlaAlaAlaSerTyrIle----- 1030  
 Db 261765 TTGAACGTAGTCACACTGCTTGAAGAACGACGTGAGTTGCTGAAATGTTGCTGGTTTT 261824  
 Qy 1031 -----GlnMetGlnTrpArgThrTyrGlnLeuGlyArg 1041  
 Db 261825 AAGCGCCATTGGTCTAATAAACTAGTCACTCCA-----AAGATTATGAATGACCAAA 261881  
 Qy 1042 IleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgMet 1061  
 Db 261882 AAACAGGAAAGTGAACAAACCCAAAGAAACACAGAACTAAAGATTGCTTTTGTGATCTA 261941  
 Qy 1062 ArgSerLysTrpSerMetLeuGlnArgLys-----GluPheGlnGlnLeuLysArg 1079  
 Db 261942 CAAGAAGCTATCAGGTTTGTGAACCTCAAAAGGACCAAGATTTAGACAA----- 261992  
 Qy 1080 AlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLysCysAla 1099  
 Db 261993 -----ATTGAAGCTAAGCAACGTAACCTTGACAACTTAGCT--GAAAAAATAATCAG 262043  
 Qy 1100 AspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTyrArgLysAlaThrIle 1119  
 Db 262044 GTCAAACTAGAACTTGATACACAGGTTTCAACGCTGCAAAACCAAAACGACACAGTA 262103  
 Qy 1120 GlnMetArgIleAsp---ArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeu 1138  
 Db 262104 CAAGCTCAGTAGAACTGGACGCTGAAACACACACAGTTAAACCTTGACCAAACTGCTTTC 262163  
 Qy 1139 GlnGlnArgLeuArgAlaIleMetLysMetArgGluGln----- 1151  
 Db 262164 AACCAAGCTAATGATCACTTTTAAACCAACGTAACCACTCACCACAAAAGATCCAAGCT 262223  
 Qy 1152 -----ArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeu 1164  
 Db 262224 TTCCACTATGAGTTGAAAAGCGTAACCAATCTTAGCTTTAAAGGGGAAAAGTTGTTT 262283  
 Qy 1165 ValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArg-AsnAlaTyrLeuArg 1184  
 Db 262284 GCAAAAGAG-----CAAGATCAACACGCAAGATCAG 262316  
 Qy 1184 gThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeuGlnMetArgArg 1204  
 Db 262317 GAGATCAACTGACGCTTTAAACACAGTTTGAAA-----AGGAATATATCATTTTGATGAA 262370  
 Qy 1204 gGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGlnIleLysPheArg 1224  
 Db 262371 GCTAAGAAAAGGGAACCTTGAAGAGCTTGAAAAGATCAGAAAGTTTAAGTCAAGC-AA 262429  
 Qy 1224 gAlaLysArgGluMetLysLysGlnArg-----AlaGluPheLeuGlnLysLysVal 1242  
 Db 262430 CGTTGATTTAGAGAGAAAAGAGAAAACCTGCTACTGATTTCACTAATTTAAATAGGT 262489  
 Qy 1242 lThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGln 1262  
 Db 262490 TCACACACACCCCAATTAACCGTGAT-----CAACTTAACAGTCAGATCAGACA 262540  
 Qy 1262 uTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSer 1282  
 Db 262541 GTTCTTATTAGAACGCCAAAACCTCCACGCTTTAGTAATGAAGCTAATGCTAAAAAGC 262600  
 Qy 1282 rMetArgPheMetArg-----AlaLysTyrArgGlyThrGlnAl 1295  
 Db 262601 CTTTTTAATTAAGCGTTTAAGAGCTTTCATCCATCTTAACCTCAAAAAGAGCGTT 262660  
 Qy 1295 aAlaValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAs 1315  
 Db 262661 AGCAATCCAAAACCTAGAGTTTGATAGCGTGATGAACACACAGAAAAAGAGTTACAGCA 262720  
 Qy 1315 nSerPheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLe 1335  
 Db 262721 AGCTACTTTTACAACTAGAACAGTTCAGTTTCAAGTTTGAAGAGCAA-----AAGTT 262765

QY 1335 uasnMetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaIleThrIleGln 1355  
Db 262766 TGAATTTGAAACAA-----CCCAACTAGTTGCTATTAA 262801  
QY 1355 nThrArgTyrArg-----AlaLysAlaMetGlnLysGlnValValLeuTy 1371  
Db 262802 AACTCAGTGTGAAAACTTAGTGATGAAAGGCACTAAACCAAGCTAGTTGAACT 262861  
QY 1371 rGlnLysGlnArgGluAlaIle-----Ly 1380  
Db 262862 AAAAAAATTATCCCAAACTATCTTCTTAATAAGATAAGGCTGAATACTCCAGCAACA 262921  
QY 1380 sValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGlu----- 1396  
Db 262922 ACTCAACAGATAACACCAATTTACTTGATCTTCAAGGAAACCTTAGAGAACCAAGA 262981  
QY 1397 ----ValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTrpTrpArgSerIle 1415  
Db 262982 TCAATTAGATAAAAAACATCGTTCTATCTTCGTTAGATTAACTAAGTTTGCAATGACTT 263041  
QY 1415 eArg-----AspMetArgLe 1420  
Db 263042 ACCTTTTGAATAAAAGCAACTGTATAAGACACGCGCATAGTTGATGATAAAACCGTCT 263101  
QY 1420 uCysLysAlaGlyTyrArgArgIleArgLeuSerSerIleGlnArgLysTrpAr 1440  
Db 263102 TTTGAAGAAATGAACGTAACTCCATTTCTTCCCAATGAACAGACGAAAA---CG 263158  
QY 1440 gAlaThrVal-----GlnAlaArgArgGlnArgGluIle 1451  
Db 263159 AGCAGTCTCGAAGTCAAAATTTCTTACTTTGAAAAACACGTAACAGCTACTGAAGC 263218  
QY 1451 ePheLeuSerThrIleArgLysValArg-----LeuLe 1462  
Db 263219 GATCCTAGCATCATATAAGAGTTAAAGAGGAGTGAACCTGCAAAAGTTACTGGT 263278  
QY 1462 tGlnAlaPheIleArgAlaThrLeu-----MetArgGlnGlnAr 1476  
Db 263279 TGAATTAGAAACAGAAAAACCAACTCAACATGATTTTGCACAAATTTCTCAAGCAACG 263338  
QY 1476 gArgGluPheGluMetLysArgArgAlaAlaValIleGlnArgArgPheArgAlaAr 1496  
Db 263339 TGAAGAGTTTGAAACCAACGCTTAAGCTCTTGGAACTGCAAAACCCCTGCAAAACCCA 263398  
QY 1496 sCys-----AlaMetLeuLysAlaArgGlnAspTyrGlnLe 1508  
Db 263399 AACTAATTCACACACTTTAAAAACCAAGCAATCCAGAGATTGAAAAACAGTTATAAAG 263458  
QY 1508 ulleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSer----- 1525  
Db 263459 GGGATGGAAGAACTTAACCTTCCAAAGAGGAGTTTGATPAAGATAAATCAAGCTTATA 263518  
QY 1526 -----MetLysGlnAlaArgGlnGlnPheValGlnLeuArgThr-----IleAlaValHi 1542  
Db 263519 TGAATACTTTAGAAAGATCGTGATGAGATTGAAGAGAAAGAGTCAGTTAAGTTAGT 263578  
QY 1542 sLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLe 1562  
Db 263579 TTTAAAGAGAGCCCAAGGAAAGCCCAACCTTTAGAAGCACAAAGCCCAACAACTTAACAT 263638  
QY 1562 uLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysAr 1582  
Db 263639 TGAATAAACACTATTGAC-----TTTAAAGAAAAAGA 263671  
QY 1582 gPheGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLy 1602  
Db 263672 GTTAAAGGCTTTTAAAGATAAGTTGATCAAGACATTGATTCAACCAAT----- 263720  
QY 1602 sValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLe 1622  
Db 263721 -----AAACACGCAAGGAGTTAAATGAGCT 263746  
QY 1622 uLeuAspIleArgLysArgIleAlaGln---LeuArgGlnGlnAlaLysAlaValAsnSe 1641

Db 263747 TTTAAATGAAACCAAGTTATTACCAATCTAATCGAAAGAGAAAGGGCTATTAAATTC 263806  
QY 1641 rValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAs 1661  
Db 263807 CAAAGATTCACTCTTAATAAGAAAGATAGAACGATTAAA---CGCCAATCCCATGATA 263863  
QY 1661 pAlaLeuAlaValLeuSerGlnLeuAspArgLeu 1672  
Db 263864 GGAGATCGGGGTGTTCGCTTGGTTGATAGATG 263897  
RESULT 7  
US-09-595-684B-30  
; Sequence 30, Application US/09595684B  
; Patent No. 6544766  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Ohashi, Cara  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Vaisberg, Eugeni  
; APPLICANT: Wood, Kenneth  
; APPLICANT: Yu, Ming  
; TITLE OF INVENTION: Human kinesins and methods of producing  
; TITLE OF INVENTION: and purifying human kinesins  
; FILE REFERENCE: cytop036  
; CURRENT APPLICATION NUMBER: US/09/595,684B  
; PRIOR FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: 09/295,612  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 8257  
; TYPE: DNA  
; ORGANISM: Human  
; US-09-595-684B-30  
Alignment Scores:  
Pred. No.: 4,76e-15 Length: 8257  
Score: 271.50 Matches: 384  
Percent Similarity: 35.55% Conservative: 365  
Best Local Similarity: 18.22% Mismatches: 857  
Query Match: 2.85% Indels: 503  
DB: Gaps: 82  
US-09-914-698-1 (1-1861) x US-09-595-684B-30 (1-8257)  
QY 8 ValLeuGluValAlaCysLysGluThrLeuGlnLeuIleAspAsnArgAsnPheArgLys 27  
Db 1093 GTTAATGAGTATCAACTGATGAAGCTCTCTG-----AAAAGGTATAGAAA 1140  
QY 28 GluValMetIleLeuLysSerLysSerAsnGlnProValLysAsnProArgLysPhe 47  
Db 1141 GAAATTAATGATCTTAAAAACAATTAGAGGAGTTTCTTTAGACGCGGGCTCAGCA 1200  
QY 48 ProThrValGlyLysThrLeuGlnLysSerProThrGlyAlaGlyLysThrMetLys 67  
Db 1201 AUGAAAAAGACCAATTGGCCCACTTTTGGAGAAAA-----GATTGCTTCAG 1251  
QY 68 SerValValSerAlaAlaValGlnGlnLysLysArgMetSerAlaAlaAlaProPro 87  
Db 1252 AAAGTACAGATGAGAAAAATTGAAAACTTTAACCGGATCTGTGTGACCTTCTTCTCCCTC 1311  
QY 88 SerLysGlnThrTrpArgValThrAlaProSerArgProAlaAlaTrpAlaHisProPro 107  
Db 1312 AGTTGCAACAGAAATAAGGCTAAAGAAACGAAGAGTACTTGTGGTC-----CCT 1365  
QY 108 ProGlnAlaProLeuValGlnLysAsnValTyrLysThrProGlnGluProValTyr 127  
Db 1366 GCGAAAAATTAAACAAATGAAGAACTCAACTATGCAGATCAATTTAATATACCAACAAT 1425  
QY 128 IleSerProGlnProArgSerLysLysGluAsnLeu----- 139

Db 1426 ATAACAACAAACACATAGCTTTCTATAAAATTTATACGAGAAATGTGATCTGTC 1485  
Qy 140 -----SerProMetThrProGlyAsnLeuLeuAspValIleAspAsnLeuArgPheThr 157  
Db 1486 TGTTCAGAGTCGTGATGTTTCAGTAACACTCTTGATACATTAAGTGAATGAATGAAT 1545  
Qy 158 ProLeuThrGluThrArgGlyGlyGlnAlaThrIlePheProAspAsnLeuAla 177  
Db 1546 CCAGCAACAAAGCTACTAAATCAG-----GAGATATAGAAGT 1584  
Qy 178 TrpProThrProThrLeuLysGlyAsnValLysSerCysAlaAsnAspMetArgProArg 197  
Db 1585 ---GAGTTGAACCTACTCTGCTGCTGACTATGATATCTGGTATTAGACTATGACAACTA 1641  
Qy 198 ArgIleThrProAspAspLeuGluAspGlnProAlaThrAsnLysThrPheAsp 215  
Db 1642 CGAAGCAAGAAAGAAAGAAATGGAATGAAATTAAGAAAGAAAGAAATGATTTGGATGAAATTT 1701  
Qy 215 ----- 215  
Db 1702 GAGGCTCTAGAAAGAAACAACTAAAGATCAAGAGATGCAACTAAATTCATGAATTTGCG 1761  
Qy 216 -----ValLysHisSerGluThrIleAsnIleSerLeuAspThrLeuAsp 230  
Db 1762 AACTTAAAGATTTAGTTAAGCATCGAAGCTATATATCAAGATCTTTGAGATGAATC 1821  
Qy 231 CysSerArgIleAsp-----GlyGlnProHisThrProLeuAsnLysThrThrIle 248  
Db 1822 AGTTCAAAAGTAGAGCTGCTTAGAGAAAGAAAGAACACCATTAAGAAAGCTACAGGAATAC 1881  
Qy 249 ValHisAlaThrHis-----ThrArgAlaLeuAlaCysIle 260  
Db 1882 ATAGACTCTCAAAAGCTAGAAATAATAAATGGAAGCTGTCATCTCAATGGAAGCAAT 1941  
Qy 261 HisGluGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeu 280  
Db 1942 GAAGACCCAAACAAATGAACAGACTCTGTTGATGCTGAACTGTAGCCCTTGATGCC 2001  
Qy 281 LysArgAspIleLysLeuValGlySerProLeuArgLysTySerGluSerMetLysAsp 300  
Db 2002 AAGAGAGAAATCAGCCCTTCTTAGAAGTCAAAATCTGGAGTTGAAGGAGAAATGAAGAA 2061  
Qy 301 LeuSerLeuLeuSerProGlnThrLysTyAlaIleGln----- 313  
Db 2062 CTTCGAACATACATACAGCAAAATGGAATTCAGATTCAGTTATATCAAGCCCAATGGAG 2121  
Qy 313 ----- 313  
Db 2122 GCMAAAAGAAATGCAAGTTGATCTGGAGAAAGAAATTAACAATCTGCTTTTAATGAGATA 2181  
Qy 313 ----- 313  
Db 2182 ACAAAACTCACCTCCTTATAGATGGCAAGTTCCAAAAGATTTGCTCTGTAATTTGGAA 2241  
Qy 314 -----GlySerMetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArg 331  
Db 2242 TTGGAAGGAAGAAATCTAGTCTTCAGAAAGAACTAAATAAGAAAGTTGAAGAAATGAA 2301  
Qy 332 TyrTyGlnGluGlnGlnIle-----GlnIleLysAla----- 342  
Db 2302 GCTTTGCGGGAAGAGTCAATTTGCTTTTCAGAAATTCAGAAATCTTACCTTCTGAGTAGAA 2361  
Qy 343 -----LysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGln 359  
Db 2362 AGGCTGAGCAAGAGATACAGACAAATCTGAAGAGCTCCATATAATACATACAGAAAAA 2421  
Qy 360 GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPheAsn-----Leu 377  
Db 2422 GATAAATGTTTT-----TCTGAAGTAGTTTCATAAGAGAGTAGAGTTCAAGGTTTACTTT 2475  
Qy 378 HisGluValGlyArgLysSerValLysGlySerProValLysAsnProHisLysArgArg 397  
Db 2476 GAAGAAATTTGGGAACAAACAAAGATGACCTAGCACTACACAGCTCGAATTTATAAGCACT 2535

Qy 398 SerHisGluLeu---SerPheSerAspAlaProSerAsn---GluSerLeuTyArgAsn 415  
Db 2536 GATCAAGAAATCCAAAATTTCAAAACCCCTTCATATGAGCTTTGAGCAAAAATATAG--- 2592  
Qy 416 GluThrValAlaIleSerProLysLysGlnArgValGluAspThrThrLeuProArg 435  
Db 2593 -----ATGGTCCTTGAGGAGAAATGAGAGAATGAATCAGGAAATAGTTAATCTCTCTAA 2646  
Qy 436 SerAlaAlaProAlaAsnAlaSer-----AlaArgSerSerSerAlaHisAlaTrp 452  
Db 2647 GAAGCCCAAAATTTGATTCGAGTTTGGTTCCTTTGAGACCGAGCTTTCTTCAAGACC 2706  
Qy 453 ProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysPro 472  
Db 2707 CAAGAACTTCAGGAGAAACACAGTGAGGTTCAAGAAAGACTTAATGAGATGAACAGCTG 2766  
Qy 473 AlaThrProArgLysValArgAspThrSerIleGln-----ProSerValLysLeuTy 490  
Db 2767 AAGGACAAATTTAGAAATAGAGATTCCTCGCTGCAAACTGTAGAAAGGGAGAAACACTG 2826  
Qy 491 AspSerGluLeuTyMetGlnThrCysIleAsnProAspProPheAlaAlaThrThr 510  
Db 2827 ATTACTGAGAACTGCAGCAAACT----- 2850  
Qy 511 IleAspProPheLeuAlaSerThrMetTyrlleuAspGluGlnAlaValAspArgHisGln 530  
Db 2851 -----TTAGAGAAATGTAATAAACTTTTAACCTCAAGAA 2880  
Qy 531 AlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAsp---LeuAspAla 549  
Db 2881 AAAGATGATCTAAACAACTCCAGAAAGCTTGCAAAATGAGAGGACCAACTCAAAAGT 2940  
Qy 550 AspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsnLysGluLeu 569  
Db 2941 GATATTCAAGTACTGTTTAACTG---AATATAGATACTCAAGAACTAATTCAGAAATGCT 2997  
Qy 570 ValValAlaProThrLysGluGlnSerMetAsnTyrlleu---ThrLysTyArgLeu 588  
Db 2998 CTTGAGTCTCTGAAACAAACATCAAGAAACAAATTAATACACTAAATCGAAATTTCTGAG 3057  
Qy 589 GluThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys 608  
Db 3058 GAAGTTTCCAGGAATTTGCATATGAGGAAATACAGGAGAAACTAAAGATGAATTTCTCAG 3117  
Qy 609 SerLysValAlaValTyValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeu 628  
Db 3118 CAAAAGATGTTGGCATAGATAAAAAACAGGATTTTGAAGCTTAAATAATACCCAAACACTA 3177  
Qy 629 HisLeuAspVal-----ValMetGlnArgThrIleLeuGluLeuLeu 642  
Db 3178 ACTGCAGATGTTAGGATAATGAGATAATGAGCAACAAAGAGATATTTCTTTAATA 3237  
Qy 643 LeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPheGlyGlyLysIle 662  
Db 3238 CAGGAGAAAATGAA-----CTCAACAATGTTAGAGAGTGTATAGCAGAAAAAGGAA 3291  
Qy 663 GlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsn----- 679  
Db 3292 CAATTGAAGACTGACCTTAAAG---GAAAATATTCGAATTCACCATTTGAAAAACAGGAGAA 3348  
Qy 680 ---ArgLeuPheArgAsnLysCysGluGlnArgTySerLysAlaTyThrLeuThr 698  
Db 3349 TTAAGACTCTTGGGATGAACCTTAAAGCAA----- 3381  
Qy 699 GluGluTyAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuPro 718  
Db 3382 CAAGAGATGTTGCACAGAAAGAACCATGCCATAAGAAAGAGAGAGGAGCTTTCTAGG 3441  
Qy 719 PheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLys 738  
Db 3442 ACCTGTGAC-----AGACTGGCAGAAAGTTGAAGAAAAACTAAAGGAAAAAG 3486



Db 5373 AAATGATGCTTAAAGACACAGGATCTGAAATAACACAGAGCACTAAGATTGCTCACAT 5432  
Qy 1389 uGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaVal----- 1405  
Db 5433 GCATCTGAAGAGCAGCAGCAAACTATTGACAAACTCAGAGCAATGTTCTGGAAGAC 5492  
Qy 1406 -----AAGATCTAAGAACTTAAGCAAACTCAACATCACTTAACTTAAAGAAAGA 5603  
Db 5493 AGATAAACTATCAATATGCAAAAGATTAGAAAATTCAAATGCTAAATACAGAA-- 5550  
Qy 1411 pTTPArgSerIleArgAspMetArgLysCysLysAlaGlyTyrArgAlaIleArgLeuSe 1431  
Db 5551 -----AAGATCTAAGAACTTAAGCAAACTCAACATCACTTAACTTAAAGAAAGA 5603  
Qy 1431 rSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArgArgGlnAluI 1451  
Db 5604 TGCAATAGACACACAGAAAAGTGTGAAATGGACCACTAAAGAAACAAATAAAGAA 5663  
Qy 1451 ePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeu 1471  
Db 5664 CCAGAGCTTAACCTGAGTAATAGAAATAGAGAAATTTAAATTTGGCTCAAGACTTCA 5723  
Qy 1471 uMetArgGlnGlnArgGluPheGluMetLysArgAlaAlaValIleGlnAr 1491  
Db 5724 TGAAGAACCTTCAAGAAATGAAATCTGTAATGAAAGAAAGAGATATCTA-----AG 5774  
Qy 1491 gArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSe 1511  
Db 5775 AAGAGTAGAG-----GAGACACTCAAACTGGAGAGAGAC---CAACTCAAGGAAAG 5822  
Qy 1511 rSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgG1 1531  
Db 5823 CCTCGAAGAACCAAGCTAGAGACTCGAATACACAGGAACTAAAGAACTGCTCGTAT 5882  
Qy 1531 nGluPheValGln-LeuArgThrIleAlaValHisLeuGlnGlnLysPheArgGly---- 1549  
Db 5883 GCTATCAAGAAAGACACAAAGAACTGTTGTAATACTTAGAGAAAAATTTCCAGAAAGAC 5942  
Qy 1550 -----LysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuL 1563  
Db 5943 AATTCAAAATTCAGACATCTCAAGAGATTTAGATAAATCAAAAG----- 5986  
Qy 1563 euArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgP 1583  
Db 5987 -----ATGATTTACAGAAAGAAAGATCCAGAACTTC---AGAAAAAGAACT 6029  
Qy 1583 heGlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLys-----ArgAlaA 1601  
Db 6030 TCAACTGCTTAGA-GTGAAGAGAGATGTCATATGAGTCATAAAAAAATTAATGAAATGG 6088  
Qy 1601 laLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArg----- 1614  
Db 6089 AACAGTTGAAGAGCAAAATTTGAGCCAACTATCTATGCAAGTGTCAGATGGATTAATCTCC 6148  
Qy 1615 -----ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuA 1633  
Db 6149 AGTTGACTAAGAACTTCAATGAAAGCTTGAAGAAATAAGA---ATTGCTAGCTAAGAAA 6205  
Qy 1633 rGlnGlnAlaLysAlaValAsnSerValArgCysLysValGlnGlnAlaValArgPheL 1653  
Db 6206 GAGATGAGCTTAAG-----AGGATAAAGAAATCTCTCAAAATGG 6244  
Qy 1653 euArgGlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArg---- 1671  
Db 6245 AAAGGACCAATTCATGCAACC-----TTAAGGGAATGATAGCTAGAGACCCAGACA 6298  
Qy 1672 -----LeuSerArgThrValProHisLeuLeuMetT 1682  
Db 6299 ACCACCAAGTAAACCTGAAAAAGGTTACTAAGTGATGGACACAGACCTTANG----- 6354  
Qy 1682 rpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerg 1702  
Db 6354 ----- 6354

Qy 1702 luValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrA 1722  
Db 6355 -----GAAAGCTGAGAGAAAGTGCTCTAGAATAAAGAGCTTTTGAAGAGTACT 6406  
Qy 1722 snSerThrThrValAsnThrPheGlnGluGlyGlyLeuValThrIleAlaGlnMetLeuL 1742  
Db 6407 CAGAG-----ATGGATGATCATTTATGAGTGTGCTTGAATAGATTGTCTC 6448  
Qy 1742 euArgTTPCysAspLysSerGluIlePheAsnThrLeuCysThrLeuIleTrpValP 1762  
Db 6449 TTGAC---TTGGAGAGGAAATTTGAATTTCCACAGAAATCATGAGAAACTGAAGTATGTGT 6505  
Qy 1762 heAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaI 1782  
Db 6506 TAACTATGTTTACAAAATAAAGAAAGAACCAACATGATGATCATATAAATTTGAAATGG 6565  
Qy 1782 leTyrMet-----ValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysG 1800  
Db 6566 ATTTTATTTGATGAAGTGGAAAAAGCAAAAGGAATTTGCTAATTAATAATACAGACCTTCAAC 6625  
Qy 1800 lnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLysIleAsn- 1819  
Db 6626 AAGATTGTGATGTACCA-----TCCAGAGAAATTAAGGATCTCAAAATTTGAACC 6673  
Qy 1820 -----PheThrProCysSerLeuProS 1827  
Db 6674 AGAATATGATGCTACATTTAGGAAATTTCTCAAGATTTCTCAGAAAGTGAATTCCTTA 6733  
Qy 1827 erLeuGluProAspPhe 1832  
Db 6734 GCATAAAGACTGAATTT 6750

## RESULT 8

US-08-328-354-5  
; Sequence 5, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8789 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single



Db 4401 CCTACTCTGGAGCTGACTTAGAGGTAGTTCACAAACAGAGAGCTACTGTTTGAAGAAAGA 4460  
Qy 681 uPheArgAsnLys 690  
Db 4461 CAATGAAATTAAGCAGAGAGTATTGTTGCTGCTTGAAGAGAACTCTCAGTGGTCACAAAG 4520  
Qy 691 699  
Db 4521 TGAGAGAAACACAGCTTCTGTGAGAGATTAGATCTATGTCAAAAAAACCACCGCAGCTGGA 4580  
Qy 699 uGluTyrAlaGluThrIleLysHisSerLeuGlnLysIleLeuPheLeuLeuProPh 719  
Db 4581 TCAGTTGCTGAAAAATGAAGAG- 4605  
Qy 719 eLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 735  
Db 4606 4652  
Qy 735 uPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerG1 755  
Db 4653 TCAGTGGCAGGAGGAGGTGAAGAAAGAACGGAACCTCTCAGACTTTGCTCTCTGA 4712  
Qy 755 uLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuG1 775  
Db 4713 TGTGAGTGAGCTGTTAAAGACAAACTCAT- 4748  
Qy 775 nHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLe 795  
Db 4749 GGAAGAGCTGAGAGTGTGGAAGAGGACTCACAGCAGCTGCTTTGACAAATGTGAGCT 4808  
Qy 795 uArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAspLeuTh 815  
Db 4809 GGAATACCAATTCACAACTGAATAAGACAGAAAGTGTGTCAGAGGAATCTGAAG 4868  
Qy 815 rArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnValLysLeuAl 835  
Db 4869 CCTGAGCCAGAGCTG- 4922  
Qy 835 aLeuGlyAlaLeu 851  
Db 4923 CTGAGGCGCAGCTGCTGGAGAAAGGTGAGTTCGCAATTGAGGCTGAGC- 4973  
Qy 851 aAlaGlnAspIleValAspGlyHisArgGluLysThrLeuLeuTyrGlnLeu1 871  
Db 4974 AACACAGGAGAGTGCATCAGCTGAGA- 5030  
Qy 871 eTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGlnLysTyrTrpPar 891  
Db 5031 TGAGCCGATGAAAGAGAGCAGCTGCACATCGCAGAGAACTGAAAGAACCGGAGCGGA 5090  
Qy 891 gArgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGluLeuMetArgAr 911  
Db 5091 GAATGATTCACTTAAGCATAAAGTTGAGAACCTTGAAGAGGAATTCAGATGTCAGAGA 5150  
Qy 911 gHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrVa 931  
Db 5151 AAACAGAGAGTGTGATCTTGATGCC- 5198  
Qy 931 lLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgAr 951  
Db 5199 AGAGACTCTAAACACAAATAGAGAGATGGCCAGAGAGCTGAAAGTTTGAATTA- 5256  
Qy 951 gTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleLeuThrIleGlnArgTr 971  
Db 5257 5277  
Qy 971 pTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAl 991  
Db 5278 5333  
Qy 991 aAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAla1 1011

Db 5334 AGACAAGTTACTCTCTTCATT- 5372  
Qy 1011 aAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnLysAlaAlaSerTyrIleG1 1031  
Db 5373 AAAGGACAGACAGATACAGATCAAGAGAATCTAAAACTGCAGTGGAGATGCTTCA 5432  
Qy 1031 nMetGlnTrpArgThrTyrGlnLeuGlyArgIle- 1042  
Db 5433 GAATCAGTTAAG- 5486  
Qy 1043 5487  
Db 5487 AATTATAGAGCCACAGACAGAGCTAGACCCACCAATAGAGGAAGAGCATCAGCTG- 5544  
Qy 1048 uArgGlnArgAspLeuIleMetPheValGlnArgMet- 1064  
Db 5545 5597  
Qy 1064 sTrpSerMetLeuGluGlnArgLysGluPheGlnGln- 1080  
Db 5598 GCTCTGTGCTTACAACTGAAGGAAGTGAAGCATCATGCAGATTTACTTTAAGGGTAG 5657  
Qy 1080 aAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCysAsnAlaAs 1100  
Db 5658 AGTGGAGAACCTTGAAGA- 5705  
Qy 1100 pTyrLeuAlaLeuArgSerSerValLeuLys- 1110  
Db 5706 GCATGCACTCTTGAGGAGAGAAATCCAAAGGAGAGGTAGAGACCCCTAAAGCAAAAT 5765  
Qy 1111 1127  
Db 5766 AGAAGGATGACCCCAAGCTCTGAGAGGTCTGGAATTAGATCTTGTACTATAAGGTGAGA 5825  
Qy 1127 sTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLy 1147  
Db 5826 AAAAGAAAATCTGACAAATGAATTACAAAAAGACAGAGCGAATATCTGAATTAGAAAT 5885  
Qy 1147 sMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLy 1167  
Db 5886 AATAAATTCATTCATTGAAAATATTTTCAAGAAAAGAG- 5930  
Qy 1167 sArgTyrArgMetArgGln- 1181  
Db 5931 GAAAGTACAGATGAAGAAAATCAAGCACTGCCATCGAGATGCTTCAACACAAATTA 5990  
Qy 1181 aTyrLeuArgThrArgLysCysIleIle- 1200  
Db 5991 AGAGTCAATGAGAGAGTGGCAGCCCTGCATAATGACCAAGAGCCTGTAAGGCCAAGA 6050  
Qy 1200 uGln- 1201  
Db 6051 GCAGAAATCTTAGTCAAGTAGAGTGTCTTGAACCTTGAGAAAGCTCAGTTGCTACAAG 6110  
Qy 1202 -MetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLys- 1220  
Db 6111 CCTTGATGAGCCAAAATAATTAATTAATTTGTTTTCATCTTCAGTGAATGCGCTCATTC 6170  
Qy 1220 nIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLy 1240  
Db 6171 AGAAGTAGAAGATGGCAAGCAGAACTGGAGAGAGGATGAAGAAATCAGTAGACTGAA 6230  
Qy 1240 sLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluAr 1260  
Db 6231 AAATCAAATTCAGACCAAGAGAGAGCTGTCTCTTAAACTGTCCAGGTGGAGAGAGCA 6290  
Qy 1260 sGln- 1276  
Db 6291 CCACTTTGGAGAGAGCAAACTTGAAGTGAAGAACTGACAGTGAATTTGGAGCAGAA 6350  
Qy 1276 sPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAl 1296  
Db 6351 G- 6392







OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10136 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-2

Alignment Scores:  
 Pred. No.: 1,15e-13 Length: 10136  
 Score: 259.00 Matches: 374  
 Percent Similarity: 34.60% Conservative: 329  
 Best Local Similarity: 18.41% Mismatches: 791  
 Query Match: 2.72% Indels: 544  
 DB: 1 Gaps: 83

US-09-914-698-1 (1-1861) x US-08-353-700-2 (1-10136)

```

QY 105 HisProProGlnAlaProLeuValGluLysAsnValTyrLys----- 119
Db 4299 CACCCTGTGCTTTGGCTCCCAATGACGAGAGTAATCTCAGCACCTTGACATGTCA 4358
QY 120 -----ThrProGlnGluProValTyrLys----- 130
Db 4359 GACAAAGAGTTCAAATGCATCTTCCGGAATTCGAAGAAATTTCTATCTTTACAAAGT 4418
QY 131 GlnProArgSerLeuLysGluAsnLeuSerProMetThrPro-----GlyAsnLeuLeu 148
Db 4419 GAACACAAAATTTTACATGATCAGCACTGTCTCAGATGAGCTCTAAATATGTCAGAGCTGCAG 4478
QY 149 AspValIleAspAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAla 168
Db 4479 ACCTATGTTGACTCATTA-----AAGCCCGAAAT 4508
QY 169 ThrIlePheProAspAsnLeuAlaIleThrProThrProThrLeuLys----- 184
Db 4509 TTGGTCTTGTCACCAATCTGAGAACTTTCAAGGTGACTTGTGAAAGGATGTCAGCTG 4568
QY 185 -----GlyAsnValLysSerCysAlaAsnAspMetArgProArgIleThr 200
Db 4569 GCGCTGGAGAGGGGCTCGTTCATCCCTGTCATCCTCTGTCCTGTCCTGCTCAGCTCTAGT 4628
QY 201 ProAspAspLeuGluAsp-----GlnProAlaThrAsnLysThrPheAspVal 216
Db 4629 CTTAGCAGTTGGGAGAGCTCCTCTTTTACAGAGCTCTTTTAGAACACAGACAGAGATATG 4688
QY 217 LysHisSerGluThrIle-----AsnIleSerLeuAspThrLeu 229
Db 4689 TCTCTTTTGAGTAATTTAGAGGGGCTGTTTCAGAAACCCAGTGCACTGTAGTAGTA 4748
QY 230 AspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThrIleVal 249
Db 4749 TTTTGCAGCAGCTCTG-----CAGACCTATGTTGACTCTATTAAAGCCGGAATTTGGTC 4802
QY 250 HisAlaThrHisThrArgAlaLeuAla-----CysIleHisGlu----- 262

```

```

Db 4803 TTGTCACGAATCTGAGAAACTTTTCAAGGTGACTTGGTGAAGGAGATGCGAGCTGGGCTTG 4862
QY 263 GluGluGlyProSerPro-----ProArgThrProThrLysSer 275
Db 4863 GAGGAGGGGCTCGTTCCATCCCTGTCATCTCTGTGTGCTGCTGACAGCTCTAGTCTTAGC 4922
QY 276 AlaIleHisAspLeuLysArgAspIleLysLeuValGlySerProLeuArgLysTyrSer 295
Db 4923 AGTTTGGGAGAC-----TCCTCCTTTTACAGAGCTCTTTT 4958
QY 296 GluSerMetLysAspLeuSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySer 315
Db 4959 GAACAGACAGAGATATGCTCTCTTTTTCAGT-----AATTAGAGAGGGGTGTTTTCAGCAAC 5015
QY 316 MetProAsnLeuAsnGluMetLysIleArgSerIleGluGlnAsnArgTyrTyrGlnGlu 335
Db 5016 CAGTGCAGTGTAGATGAAGTATTTTGGCAGCTCTGCGAG----- 5057
QY 336 GlnGlnIleGlnIleLysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeu 355
Db 5058 -----GAGAACTGACACGAGAAAGAAACCCCTTCGCGCCCGCAGCG 5096
QY 356 AlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe 375
Db 5097 AAGGTGTGTTGAAGACTT-----GAGTCCCTCTGTGAGGTGTACCGGAG 5141
QY 376 AsnLeuHisGluValGlyArgLysSerValLysGlySerProValLysAsnProHisLys 395
Db 5142 TCCCTCGAAGCTAGAGAGAAATGGAAGTCAAGGATATGAAATAAGGAAATT 5201
QY 396 ArgArgSerHisGluLeuSerPheSer-----AspAlaProSerAsnGlu 410
Db 5202 CAAGAGCTCGAGCAGTATTAAAGTTCTGAAAGGCAAGAGCTTGACTGCTTAGGAAGCAG 5261
QY 411 SerLeuTyrArgAsnGluThrValAlaIleSerProProLys----- 424
Db 5262 TATTGTGCAAAAATGA-ACAGTGGCAACAGAAAGCTGACAGCGTACTCTGGAATGA 5320
QY 425 -----LysGlnArgValGluAspThrThrLeuPro-----ArgSer 436
Db 5321 GTCCAAGTTGGCGGAGAAAGAAACAGACGGAACACTGTCACCTGAGCTGGAAGTAGC 5380
QY 437 AlaAlaProAlaAsnAlaSerAlaArgSerSerAlaHisAlaIleThrProHisAlaGln 456
Db 5381 ACGACTCCAGCTACAAAGTCTGGACTTAAGTCTCGGTCTTCTGCTTGCA-TCGACACAG 5439
QY 457 SerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArg 476
Db 5440 AAGATGCTATTTCAGGCGCAATGAGAGCTGTGAC-ATATCAAAAGAACATATCTTCAGAA 5498
QY 477 LysValArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMet 496
Db 5499 ACTCAGAAAGAACCAACCAAGCATGATGTTTCATCAGATTTGTGATAAAGATGCTCAGCAG 5558
QY 497 GlnThrCysIleAspProAspProPheAlaAlaThrThrThrIleAspProPhe----- 514
Db 5559 GACCTCAATCTAGACATTGAGAAATAACTGAGACTGTGTGTCAGTGAACCCAGCAGAGAG 5618
QY 515 -----LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529
Db 5619 TGCTCTGGGGAACAGTCCCGACATACCAATTAT---GAGCCTCCAGGGGGAAGATAAACCC 5675
QY 530 GlnAlaAspPheLys-----LysTrpLeuAsnAlaLeuValSer 542
Db 5676 CAGGGCTCTTCAGAAATGCAATTTCTGAAATGCTCATTTTCTGGTCTCTTAATGCTTTGTA 5732
QY 543 IleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsn 562
Db 5733 ---CCTATCGATTTC-----CTGGGGAATCAGGAAGATATCCATATCTTCAACTG 5780
QY 563 GluVal----- 564
Db 5781 CCGGTAAGAAAGACATCAAAATGAGAAATTTGAGATTACTTTCATGTGATGAGGACCGTGAC 5840

```

QY	565	ArgAsnLysGluLeuValValAlaProThrLysGluGluGlnSerMetAsnTyrLeuThr	584
DB	5841	AGAAAAGTTGAAAGTTGCTAAATCAAAATGAAGAAATTAGACTCAAAACTCCATTACAG	5900
QY	585	LysTyrArgLeuGluThrLeuArgLysAlaValGluLeu	598
DB	5901	GAGGTACAACTAATGACCAAAATTGAAGCATGCTAGAAATGGAAAAAATAGTTGGGAA	5960
QY	599	-----PhePheSerGluGlnMetArg-----LeuProCysSerLysVal	611
DB	5961	CTTAAGAAAGAAAAGCTCAGATTTAAGTGAATAATGGAAATATTTCTTGGATCACCAG	6020
QY	612	AlaValTyrValAsnLysGlnAlaLeuArg---IleArgSerAspArgAsnLeuHisLeu	630
DB	6021	GAGTTACTCCAGAGAGTAGAAACTTCTGAAGGCCTCAATTCCTGANTTAGAAATGCATGCCA	6080
QY	631	AspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTyrLeu	650
DB	6081	GATAAATCATCAGCTGAAGATATTGGAGATAATGTGCCAAGTGAATGACAGCTGGAAG	6140
QY	651	ArgLeuGlyLeuGluVal-----Val	657
DB	6141	GAGAGATTTCTTGATGTGGAATAATGAGCTGAGTAGGATCAGATCGGAGAAGCTAGCATT	6200
QY	658	PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle	677
DB	6201	GAGCATGAAGCCCTTACTCGAGGCTGACTTAGAGGTATTCAAACAGAGAAGCTATGT	6260
QY	678	LeuAsnArgLeuPheArgAsnLys-----CysGluGluGlnArg-----	690
DB	6261	TTAGAAAAAGCAATGAAATAACACAGAGGTATTGTCTGCTTGAAGAAGAACTCTCA	6320
QY	691	LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro-----	733
DB	6321	GTGCTCAAAAGTCAGAGAAACCAGCTTCGTGGAGAAATTAGATATACTATGTCAAAAAACC	6380
QY	696	ThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPhe	715
DB	6381	ACGCATGAGCTAGTTGTTCTGAAAAAATGAAGAG-----	6416
QY	716	LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro-----	733
DB	6417	-----AAAACACAGAGCTTCAGTCTCATCAAAGTGTGT	6452
QY	734	-----CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArg	751
DB	6453	CTCATTTGCATTCAGGTGGCAGAGGAGAGGTGAAGGAAAGACGGAACCTCTTCAGACT	6512
QY	752	PheSerSerGluLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLeuGly	771
DB	6513	TTTGCTCTGATGTGAGTGAGCTGTTAAAGACAAACTCAT-----	6554
QY	772	TyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnLeu	791
DB	6555	-----CTCAGAAAAGCTGCAGAGTTTGAAAAGAGACTCACAGGCACCTGCTTTTGACA	6608
QY	792	AlaValAspLeuArgAspGlyValArgLeuThrArgValGluValIleLeuLeuArg	811
DB	6609	AAATGTGAGCTGAAACCAAAATTTGCACAACCTGAATAAAGAGAAAAGAAATTTGTTGCAAG	6668
QY	812	AspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsn	831
DB	6669	GAATCTGAAAGCTGCAGGCCAGACTG---AGTGAATCAGATTATGAAGAAGCTG---AAT	6722
QY	832	ValLysLeuAlaLeuGlyAlaLeu-----GlyGluAlaAsnPheGlnLeuGly	847
DB	6723	GTCTCCAGGCCCTTGGAGGCCGCGACTGGTGGAAAAGGTGAGTTGCAATTTGAGGTGAGC	6782
QY	848	GlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeu	867
DB	6783	-----TCAACACAGAGGAAGTGCATCAGCTCAGA---AGAGGCATCAGAAACTGT	6830

QY	868	TpGlnLeuIleTyTrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln	887
Db	6831	AGAGTTTCGCAITTAGCCCGATGAAGAAGCAGCTGCACATCGCAGAGAACCTGAAGAAG	6890
QY	888	LysTTPtTpArgArgHisTfTpLeuHisvalValledGlnnArgArgileAArgHsiLysGlu	907
Db	6891	COCGACGGGAGAAATGATTCTTAAGATAAAGATTGAGAACCCTTGAAAAGGGAATGCCAG	6950
QY	908	LeuMetArgArgHisArgAlaAlaThrValiledGlnAlaValPheArgGlyHisGlnMet	927
Db	6951	ANGTCAGAAGAAACCCAGACTAGTGATCTTGATGCC-----GAGAATTC	6998
QY	928	ArgHystfYrVallLysLeuPheLysThrgluargThrGlnAlaAlaillelLeuGlnLys	947
Db	6999	AAAGCAGAAGTAGAGACTCTTAANAACAATAAGAGAGATGCCAGAACCTTGAGATT	7058
QY	948	PheThrArgArgTyTrLeuAlaGlnLysGlnLeuTyTrGlnSerTyrHisserillelThr	967
Db	7059	TTTTGAATTA-----GACCTTGTGCAG	7079
QY	968	IleGlnArgTgTtpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnargPheValGlu	987
Db	7080	TTAGTGCT-----GAAAAAGAAATCTGCAAAACAATAACAAGAAAAACAAGCTCAC	7133
QY	988	LeuArgGluAlaAlaIlePheLeuGlnArgileTtpArgArgArgLeupHealLysLys	1007
Db	7134	TTGTCAAGACTAGACAAGTTACTCTCTTCATTT-----AAAGT	7172
QY	1008	LeuLeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnAlaAla	1027
Db	7173	CTGTTAAGAAAGAGGACGAGACAGATACAGATCAAAGAGAATCTAAAACTGCAGTG	7232
QY	1028	SerTyrileGlnMetGlnTtpArgThrTyTrGlnLeuGlyArgile-----	1042
Db	7233	GAGATGCTTCAGAAATCATGTTAAG-----GAGCTAAATGAGCGACTAGCAGCCTTGCT	7286
QY	1043	-----GlnArg	1044
Db	7287	GGTGACCAAGAAATTATGAAGGCCACAGAACAGCTCTAGACCCCAATAGAGNAGAG	7346
QY	1045	HisGluPheLeuArgGlnArgAspLeulleMetPheValGlnArgMet-----	1061
Db	7347	CATCAGCTG-----AGAAATAGCAITGAAAAGCTGAGAGCCGCTAGAAAGCTGAT	7397
QY	1062	--ArgSerLysTtpSerMetIleuGluGlnArgLysGluPheGlnGln-----	1076
Db	7398	GAAAGAGCAGCTGTGTCTTTACAACTGAGGAAAGTGGCATCATCAGATTTA	7457
QY	1077	LeuLysArgAlaAlaIleAsnIleGlnArgTtpArgAlaLysLeuSerMetArgLys	1096
Db	7458	CTTAAGGTTAGTGGAGAACCTTGAAAGA-----GAGCTAGATAGCCAGG	7505
QY	1097	CysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----	1110
Db	7506	ACAAACCAAGCATGCTCTTTGAGGCAGAGAAATTCCAAAGGAGAGGTAGAGACCCTA	7565
QY	1111	-----ValGlnAlaTyArgLysAlaThrIleGlnMetArgile	1123
Db	7566	AAGCAAAATAGAGGGATGACCCAAGCTCTGAGAGGTCTGGAATTTAGATGTTTACT	7625
QY	1124	AspArgAsnHisTfYrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArg	1143
Db	7626	ATAAGTTCAGAAAAGAAATCTGCAATGAATTCAAAAGACAGACGGAATATCT	7685
QY	1144	AlaIleMetLysMetArgGluGlnArgGluAsentYrLeuArgLeuArgAsnAlaserile	1163
Db	7686	GAATTAGAANTATAAATTCATCATTTGAAAATATTTTGAAGAAAAAGAG-----	7736
QY	1164	LeuValGlnLysargTyTrArgMetArgGln-----GlnMetIlegln	1177
Db	7737	-----CAAGAGAAAGTCAAGATGAAGAAAAAATCAAGCCTGCGATGAGATGCTTCAA	7790
QY	1178	AspArgAsnAlaTyTrLeuArgThrArgLysCysIlelle-----AsnValIcLnArgArqTfp	1196

Db 7791 ACACAAATTAAGAGAGCTCAATGAGAGAGTGGCAGCCCTGCATAATGACCAAGAGCCTGT 7850  
 Qy 1197 ArgAlaThrLeuGln-  
 Db 7851 AAGCCAAAGAGCAGAAATCTTAGTAGTCAAGTAGAGTGTCTTGAACCTTGAGAAGCCTCAG 7910  
 Qy 1202 -----MetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLys 1217  
 Db 7911 TTGCTACAGCCCTTGATGAGCCCAAAATAATATATTTTTCGAATCTTCAGTGAAA 7970  
 Qy 1218 ---ArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPhe 1236  
 Db 7971 GGCCTCAITTCAGAAAGTAGAAGTGCAGCAGCAAACTGGAGAAAGAGTGAAGAAATC 8030  
 Qy 1237 LeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMet 1256  
 Db 8031 AGTAGACTGAAAAATCAAATTCAGACCAAGCAGCAGCTGTCTCTAAACTGTCCAGGTG 8090  
 Qy 1257 ArgLysGluArgGln-----GluTyrLeuHisLeuArgGluValThrLys 1272  
 Db 8091 GAAGAGAGCACCACCTTTGGAGGAGCAGCAAACTTAGAACTGAGAAATCTGACGGTGGAA 8150  
 Qy 1273 LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly 1292  
 Db 8151 TTGGAGCAGAG-----ATCCAAGTGTCTACAAATCCAAAATCGCTCT 8192  
 Qy 1293 ThrGlnAlaValSerCysLeuGlnMetHisTrpArgAsn-----HisLeu 1308  
 Db 8193 TTGCAGGACACATTAAGAAGTGTGCAGAGTCTTTTCAAGAAATCTAGAGAAATGAGCTTCAA 8252  
 Qy 1309 LeuArgLysArgGluArgAsnSerPheLeu----- 1318  
 Db 8253 TTGACAAAATGGCAAAATGCTTTTGTGAAAAGTAAACAAAATGACTCCAAAGGAA 8312  
 Qy 1319 -----GlnLeuArgGlnAlaAlaIleThrLeuGlnArg 1330  
 Db 8313 ACTGAGCTGCAGAGGAAATGCATGAGATGGCAGCAAGAAACAGCAGAGCTGCAAGAGAA 8372  
 Qy 1331 TyrArgAlaArgLeuAsn-----MetIleLysGlnLeuLys 1342  
 Db 8373 CTCAGTGGAGAGAAAATAGGCTAGCTGGAGAGTTGCAGATTACTGTTGGAAGAAATAAG 8432  
 Qy 1343 SerTyrAla---GlnLeuLysGlnAlaAlaIleThrLeuGlnThrArgTyrArgAlaLys 1361  
 Db 8433 ACCACCAAGATCAATTAAGAGAGCTCACACTAGAAAATAGTGAATTTGAAGAGAGCCTA 8492  
 Qy 1362 LysAlaMetGlnLys---GlnValValLeuTyrGlnLysGlnArgGluAlaIleLys 1380  
 Db 8493 GATTGCATGCACAAAGACAGCTGGAAAAGGAGGAAAGTGAGAGAGGAAATAGCTGAA 8552  
 Qy 1381 ValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLys 1400  
 Db 8553 TATCAGCTACGGCTTCAT---GAAGCTGAAAAGAACACACAGGCTTTGCTTTTGACACA 8609  
 Qy 1401 GlnArgGlnAlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeu 1420  
 Db 8610 AACAAACAGTATGAAGTAGAATCCAGACATACCGAGAAATTTGACTTCTTAAGAGAA 8669  
 Qy 1421 CysLysAlaGlyTyrArg-----ArgIleArgLeuSerSerLeuSerIleGln 1436  
 Db 8670 TGCTCAGTTCACAGAGCTGGAGATAGACCTTTTAAAGCTTAGTAAAGAGAGCTCAAT 8729  
 Qy 1437 ArgLysTrpArgAlaThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIle 1456  
 Db 8730 AATTCATTGAAGAGCTACTACTCAG-----ATTTTGAAGAAATG 8768  
 Qy 1457 ArgLysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArg 1476  
 Db 8769 AAGAAAACCAAGATGGACAAATCTAAATATGTAATTCAGTTGAAGAGGAAATGAACGT 8828  
 Qy 1477 ArgGluPheGluMetLysArgAlaAlaValIleGlnArgPheArgAlaArg 1496

Db 8829 GCCCAGGCGAAAAATGAAG-----TTGTTGATCAAA-----TCC 8861  
 Qy 1497 CysAlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuVal 1516  
 Db 8862 TGTAACAGCTGGAA-----GAGGAAAAGAGATACTGCGAAGAAAGACTC----- 8906  
 Qy 1517 GlnArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeu 1536  
 Db 8907 -----TCTCAACTTCAAGCTGCACAGAG----- 8930  
 Qy 1537 ArgThrIleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIleGln 1556  
 Db 8931 -----AAGCAGAAA----- 8939  
 Qy 1557 ArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAlaArgGly 1576  
 Db 8940 ----- 8945  
 Qy 1577 PheMetAlaArgLysArgPheGlnAlaLeuMetThr-----ProGluMetMetAspLeu 1594  
 Db 8946 ACTGTATGGATACCAAGTCGATGAATTAACAACCTGAGATCAAGAACTGAAGAAACT 9005  
 Qy 1595 IleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArg 1614  
 Db 9006 CTTTGAAGAAAA-----ACCAAGGAGCAGATGAATAC----- 9038  
 Qy 1615 ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGln 1634  
 Db 9039 ---TTGTAATAGTACTGTCTCTTATAGCCATGAAGATTTAGAGAAAGCTAAAGAG 9095  
 Qy 1635 GluAlaLysAlaValAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArg 1654  
 Db 9096 ATGTTAGAGACACAGTGGCCCTCTGTGTTCACAGCAATCTAA-ACAAGATTTCCGAGG 9154  
 Qy 1655 GlyArgPheIleAlaSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArg 1674  
 Db 9155 GTCTCTTTGTAGTCCAGTTGT-----TCCAGG 9184  
 Qy 1675 ThrValProHis-----LeuLeuMetTrp----- 1682  
 Db 9185 ACCATCTCCAATCCCTTCTGTACTGAAAAGAGGTTATCATCTGGCCAAAATAAGCTTC 9244  
 Qy 1683 CysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGlu 1702  
 Db 9245 AGCCAGAGGCAAGATCCAGTGGAA-TATGGGAGAATGCTAGAG----- 9288  
 Qy 1703 ValAspLysGlnLeuIleGluArgCysSerArgIleIleLeuAsnLeuAlaArgTyrAsn 1722  
 Db 9289 ---GACCAACACCTGCTACCCAG-----AGAGCTTTTCTAAAAAAGCAAGAAAGCAG 9339  
 Qy 1723 SerThrThrValAsnThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeu 1742  
 Db 9340 TCATGAGTGGTATTACCCCTGCAGAA-----ACACGGAAGGTACTG 9381  
 Qy 1743 ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpValPhe 1762  
 Db 9382 AGTTTGACCCAGAGGAC-----TTCAGAGATTG----- 9411  
 Qy 1763 AlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAlaIle 1782  
 Db 9412 -----TAAAGAAAGGG-TTGTGTGACATCCGACAGGAAAGAGTAGCCCA 9455  
 Qy 1783 TyrMetValArgGluThr-----LysLysLeuValAlaArgLys 1795  
 Db 9456 TATATCTGCGAAGAACCAACCATCGCAACTCGGACCCCGCTCGCTGCACAGAG 9515  
 Qy 1796 GluLysMet-----LysGlnAsnAlaArgLysProProMetThr 1809  
 Db 9516 TTAGCGCTATCCCACTGAGTCTCGGCAAGAAATCTTGGCAGAGCTCTCCAAACCAACA 9575  
 Qy 1810 SerGlyArgTyrLysSerGlnLysIleAsnPheThrProCysSerSerLeuProSerLeuGlu 1829  
 Db 9576 GCTGGTGCAGCAGATCATCAAAAGTCAAACTGTGCTCAGCGAGC-----CCA 9623



Db 5440 AAGATGCTATTACAGGCCGAAATGAGAGCTGTGAC-ATATCAAAAGAAACATATCTTCAGAA 5498  
Qy 477 LysValArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMet 496  
Db 5499 ACTACAGAAAGACCAACCAAGCATGATGTTTCATCAGATTTGTGATAAAGATGCTCAGCAG 5558  
Qy 497 GlnThrCysIleAsnProAspPheAlaAlaThrThrIleAspProphe 514  
Db 5559 GACCTCAATCTAGACATTGAGAAATAACTGAGACTGGTGCAGTGAACCCACAGGAGAG 5618  
Qy 515 -----LeuAlaSerThrMetTyrLeuAspGluGlnAlaValAspArgHis 529  
Db 5619 TGCTCTGGGGAACAGCTCCGCCAGATACCAATTAT---GAGCCTCCAGGGGGAATAAAACC 5675  
Qy 530 GlnAlaAspPheLys-----LysTyrLeuAsnAlaLeuValSer 542  
Db 5676 CAGGCTCTTCAGAAATGATTTCTGAATGTCTATTTCTGGTCTTAATGCTTTGGTA--- 5732  
Qy 543 IleProAlaAspLeuAspAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsn 562  
Db 5733 ---CCTATGGATTTC-----CTGGGGAATCAGGAAGATATCCATAATCTTCAACTG 5780  
Qy 563 GluVal----- 564  
Db 5781 CGGGTAAAGAGACATCAAAATGAGAAATTTGAGATTACTTATGTGATAGAGCGGTGAC 5840  
Qy 565 ArgAsnLysGluLeuValValAlaProThrLysGluGlnGlnSerMetAsnTyrIleuThr 584  
Db 5841 AGAAAGTTGAAAGTTTCTTAATGAATGAAGAAATTTAGACTCAAACTCCATTTACAG 5900  
Qy 595 LysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLeu----- 598  
Db 5901 GAGGTACAACTAATGACCAAAATTTGAAGCATGATAGAAATTTGAAAAATAGTTGGGAA 5960  
Qy 599 -----PhePheSerGluGlnMetArg---LeuProCysSerLysVal 611  
Db 5961 CTTAAGAAAGAAACTCAGATTTAAGTGAATAATTTGAATATTTTCTGTGATCACCAG 6020  
Qy 612 AlaValTyrValAsnLysGlnAlaLeuArg---IleArgSerAspArgAsnLeuHisLeu 630  
Db 6021 GAGTTACTCCAGAGATGAAACTTCTGAAGGCCTCAATCTGATTTAGAAATGATGATCA 6080  
Qy 631 AspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTripleu 650  
Db 6081 GATAAATCATCAGCTGAAGATATTGGAGATATTGGCCCAAGGTCAATGACAGCTGGAG 6140  
Qy 651 ArgLeuGlyLeuGluVal-----Val 657  
Db 6141 GAGAGATTTCTTGATGTGAAATAGCTGAGTAGATCAGATCGGAGAAAGCTAGCAATT 6200  
Qy 658 PheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIle 677  
Db 6201 GAGCATGAGCCCTCTACCTGGAGCTGACTTAGAGGTAGTTTCAACACAGAAAGCTATGT 6260  
Qy 678 LeuAsnArgLeuPheArgAsnLys-----CysGluGluGlnArg----- 690  
Db 6261 TTAGAAAAGAGCAATGAAATAAGCAGAGAGTATTGTCTGCTGCTTGAAGAAAGCTCTCA 6320  
Qy 691 -----TyrSerLysAlaTyr 695  
Db 6321 GTGCTCAAGTGAAGAAACCCAGCTTCGTGGAGAAATTTAGATACTATGTCAAAAAAACC 6380  
Qy 696 ThrLeuThrGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPhe 715  
Db 6381 ACGGACTGGATCAGTTCTGTGAAAAATGAGGAG----- 6416  
Qy 716 LeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro----- 733  
Db 6417 -----AAACACAAGAGCTTGTAGTCTCATCAAGTGAAGTGT 6452  
Qy 734 -----CysLeuPheValLysSerProHisLysGluThrLysAspIleLeuLeuArg 751  
Db 6453 CTCATTGTCATTGAGTGGCAGAGCAGAGTGAAGGAAAGACGGAACCTCTTCAGACT 6512

Qy 752 PheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGly 771  
Db 6513 TTGTCTCTTGATGTGAGTGAAGTGTGTTAAAGACAAAACACTCAT----- 6554  
Qy 772 TyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeu 791  
Db 6555 -----CTCCAGGAAAGCTGCAGAGTTTGGAAAGAGACTCACGGCATGCTCTTTGACA 6608  
Qy 792 AlaValAspLeuArgAspGlyValArgValArgLeuThrArgValGluValIleLeuLeuArg 811  
Db 6609 AAATGTGAGCTGGAACCAACCAATTCACAACTGAATAAAGAGAGAAATGCTTGTCTCAAG 6668  
Qy 812 AspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsn 831  
Db 6669 GAATCTGAAGCCCTGCAGCCAGACTG---AGTGAATCAGATTATGAAAAGCTG---AAT 6722  
Qy 832 VallysLeuAlaLeuGlyAlaLeu-----GlyGluAlaAsnPheGlnLeuGly 847  
Db 6723 CTCTCCAAGGCTTGGAGGCGCACTGGTGAGAAAGGTGAGTTGCGATTGAGGCTGAGC 6782  
Qy 848 GlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeu 867  
Db 6783 -----TCAACACACAGGAGGAAGTGCATCAGCTGAGA---AGAGGCATCGAGAACTG 6830  
Qy 868 TrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGln 887  
Db 6831 AGAGTTCGATTCGAGGCGGATGAAAGAGCAGCTGCATCGCAGAGAAATGAAAGAA 6890  
Qy 888 LysTrpTrpArgHisTrpLeuHisValIleGlnArgArgIleArgHisLysGlu 907  
Db 6891 CGCGAGCGGAGATGATTCTACTTAAGATTAAGTTGAGAACCTTGAAGGGAATTCAG 6950  
Qy 908 LeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMet 927  
Db 6951 ATGTCCAGAAAGAACCCAGAGCTAGTGTATTTCTGATGCC-----GAGAATTC 6998  
Qy 928 ArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947  
Db 6999 AAACAGAGTAGAGACTCTAAACACACAAATGAAGAGAGATGGCCAGAGCCTGAAGATT 7058  
Qy 948 PheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThr 967  
Db 7059 TTTGAAATTA-----GACCTTGTCTACG 7079  
Qy 968 IleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGlu 987  
Db 7080 TTAAGGTCT-----GAAAAAGAAAAATCTGACAAAAACAAATACAGAAAAACAAGGTCTCAG 7133  
Qy 988 LeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLys 1007  
Db 7134 TTGCAGAACTAGACAAAGTTACTCTCTTCATTT-----AAAAAGT 7172  
Qy 1008 LeuLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAla 1027  
Db 7173 CTGTTAGAGAAAGGAGGACGACAGATACAGATCAAGAGAGATCTAAACCTGCATG 7232  
Qy 1028 SerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIle----- 1042  
Db 7233 CAGATGCTTCAGAAATCAGTTAAAG-----CAGCTAAATGAGGCAGTAGCAGCCTTGCT 7286  
Qy 1043 -----GlnArg 1044  
Db 7287 GGTGACCAAGAAATTTAAGAGCCACAGACAGAGTCTAGACCCACCAATAGAGGAGAG 7346  
Qy 1045 HisGluPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMet----- 1061  
Db 7347 CATCAGCTG-----AGAAATAGCATTTCAAAGCTGAGAGCCCGCTAGAGAGTGT 7397  
Qy 1062 ---ArgSerLysTrpSerMetLeuGluGlnArgLysGluPheGlnGln----- 1076  
Db 7398 GAAAAGAGCAGCTCTGCTCTTACAACTGAAGGAGAAAGTGAAGCATCATGTCAGATTTA 7457

QY	1077	LeuLysArgAlaalaIleAenlleGlnGlnArgTyrPArgAlaLysIeuSerMetArgLys	1096
Db	7458	CTTTAAAGGGTAGAGTGGAAGACCTTTGAAAGA-----GAGCTAGAGATAGCCAGG	7505
QY	1097	CysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----	1110
Db	7506	ACAACCAAGCAGCATGCAGCTCTTGAGCGAGAGATTCCAAGAGAGGTAGAGACCCTTA	7565
QY	1111	-----ValGlnAlaTyrArgLysAlaThrIleGlnMetArgIle	1123
Db	7566	AAAGCAAAAATAAGAGGGATGACCCAAAGCTCGAGAGGTCGGAATTAGATGTTGTACT	7625
QY	1124	AspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArg	1143
Db	7626	ATAAGGTCAGAAAAAGAAATCTGC AAAATGAATTA CAAAAGACAGACGACCAATCT	7685
QY	1144	AlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIle	1163
Db	7686	GAATTAGAAATAATAATTCAATCATTTGAAATAATTTTGC AAGAAAAAGAG-----	7736
QY	1164	LeuValGlnLysArgTyrArgMetArgGln-----GlnMetIleGln	1177
Db	7737	-----CAAAGAAAAGTACAGATGAAGAAAAATAACAGCACTGCCATGGAGATGCTTCAA	7790
QY	1178	AspArgAsnAlaTyrLeuAlaTyrArgLysCysIleIle--AsnValGlnArgArgTyr	1196
Db	7791	ACACAATTAAGAGCTCAATGAGAGAGTGGCAGCCCTGCATATGACCAAGAGCTGT	7850
QY	1197	ArgAlaThrLeuGln-----	1201
Db	7851	AAGGCCAAAGACAGAGAACTTTAGTAGTCAAGTAGAGTGTCTTGAACTTGAGAAGGCTCAG	7910
QY	1202	-----MetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLys	1217
Db	7911	TTGGTCAACAGCCCTTGATGAGGCCCAAAATAATTATATGTGTTTGC AATCTTCAGTGAAA	7970
QY	1218	---ArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPhe	1236
Db	7971	GGCCTCATTCAGAAAGTAGAAGATGGCAAGCAGAAACTGCGAAGAAGGATGAAGAAATC	8030
QY	1237	LeuGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMet	1256
Db	8031	AGTAGACTGAAAATAATCAAAATTC AAGACCAAGACAGAGCTGTGCTCTAACTGTCCCAGGT	8090
QY	1257	ArgLysGluArgGln-----GlnTyrLeuHisLeuArgGluValThrIleLys	1272
Db	8091	GAAAGAGAGCACCAACTTTGGAAGGACCAAACTTAGAACCTGAGAAATCTGACGGTGAA	8150
QY	1273	LeuGlnArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGly	1292
Db	8151	TTGAGCAGAGAAG-----ATCCAAGTGTACAATCCAAAATGCGCTCT	8192
QY	1293	ThrGlnAlaAlaValSerCysLeuGlnMetHisTyrArgAsn-----HisLeu	1308
Db	8193	TTGCAGACACATTAGAAAGTCTGCAGAGTCTTAC AAGATCTAGAGAATGAGCTTGAA	8252
QY	1309	LeuArgLysArgGluArgAsnSerPheLeu-----	1318
Db	8253	TTGCAAAAATGGACAAAATGCTCTTTGTTGAAAAAGTAAACAAAAATGACTGCAAAAGGAA	8312
QY	1319	-----GlnLeuArgGlnAlaAlaIleThrLeuGlnArgArg	1330
Db	8313	ACTGAGCTGCAGAGGAAATGCATGATGATGCGACAGAAAACAGCAGAGCTGCAGAAGAA	8372
QY	1331	TyrArgAlaArgLeuAsn-----MetIleLysGlnLeuLys	1342
Db	8373	CTCAGTGGAGAGAAAAATAGGCTAGCTGGAGAGTTGCAGTTACTGTTGG AAGAAATAAAG	8432
QY	1343	SerTyrAla---GlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLys	1361
Db	8433	AGCAAGCAAGATCAATTGAAGAGCTCACACTAGAAAATAGTGAATTTGAAGAAGAGCTTA	8492
QY	1362	LysAlaMetGlnLys---GlnValValLeuTyrGlnLysGlnArgGluAlaIleLys	1380

[illegible]





Db 1348 CTCGGTTACCACTCCGTTACCAACCGCGCTCCACCTCCAGCTCAAGCTGATGACAGATG 1289  
Qy 1139 GlnGlnArgLeuArgAlaIleMetLysMetArgGlnArgGluAsnTyrLeuArg--- 1157  
Db 1288 CAGATG---CTGAAGCAGATCTGAGCTTTCAATAGACCTGCAAGATCTTTTAAGACAT 1232  
Qy 1158 -----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMet 1175  
Db 1231 CCTCAAAATCGTCAATTGAGATTGTGGCTTGCTACTGCTGATGCTTTAGCAGAGGCTT 1172  
Qy 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle----- 1190  
Db 1171 TTGCAGAGGCTCCATTACTCTCGAAGAGTCTTTAATGCTCTCTTAATGCGCTCCCTTATAC 1112  
Qy 1191 -----AsnValGlnArgArgTyrArgAlaThrLeuGlnMet 1202  
Db 1111 CACTCATTGCTAGAGCTGAACGATGACGAGCGGCTGCGGCGAGCGGACGAC 1052  
Qy 1203 ArgArgGluArg-----LysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln 1220  
Db 1051 CAGCGCTCGCGCTCTCTCTCCACACCTCCACCTCTCGCGCACACCGAGCGCTG--- 995  
Qy 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu-----Phe 1236  
Db 994 -----CGCGCGAGCTGACGAGCGGCGAGCGAGCGAGCTGCGGCGCTC 950  
Qy 1237 LeuGlnLeuLysLysValThrLeuValGlnLysArgArg-----AlaLeuLeu 1254  
Db 949 CTCTTCTTCACCGCGTCTACTCTCTCTGTCACCGTAGAACAACACACCGAGTCTCTC 890  
Qy 1255 GlnMetArgLysGluArgGlnGlnTyrLeuHisLeuArgGluValThrLysLeuGln 1274  
Db 889 CAGCTCGCGCAGCAGCGCGCGCAGCAGCAGCGCGCAGCAGATCTCTCCATATCTC 830  
Qy 1275 ArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGln 1294  
Db 829 CGA-----GGCCACCAAGACCTCCGAGGCTCCAACTGCCACCGAGTCTCTCAA 779  
Qy 1295 AlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArg 1314  
Db 778 GTCCGCCACCGATCTCTCAAGTCCACCTCGC-----CTCTGACGCGCGCAGCGG 725  
Qy 1315 AsnSerPheLeuGlnLeuArgGln-----AlaAlaIleThrLeuGln 1328  
Db 724 -----CTGGCGTGCAGCAGCAGCTCTCCAACTGCTGCGCTCTCTCTGCGAG 674  
Qy 1329 ArgArgTyrArgAlaArgLeu-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
Db 673 CGCAGCAGCGGTGCGGTGCGGAGCAGCTCCACCACTCCACCCCGCGCGCAGCAG 614  
Qy 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
Db 613 CGGCAG----- 608  
Qy 1368 valValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgTyrArgGly 1387  
Db 607 -----CTCAGCAGCAGCAGCAGCAGC-----CTCCAGCTCCACCGCTCCACCGAGG 560  
Qy 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
Db 559 CAA-----GAGCTGCACCAAGATCATCGAAAT---CTAAATCAA 524  
Qy 1408 LeuGlnLysTyrTrpArgSerIleArg---AspMetArgLeuCysLysAlaGlyTyrArg 1426  
Db 523 ATAAATCGTCTCCACCGCCCAAGTCCGCCAAGACCTCCAAGTC-----CGC 479  
Qy 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArg 1446  
Db 478 CAACTCCACCTCCGAGACCTC-----CGGCCAATCCAAAGTCCGA 440  
Qy 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466  
Db 439 GTCCAGCACTGCTCTGCTC-----CTGGCGCTGCTCCAGCAGCAGCAGCAGC 395

Qy 1467 ArgAlaThrLeuLeuMetArgGlnGln-----ArgArgGluPheGluMetLysArg 1483  
Db 394 CAGCACCACCTCCGCGCACCAACACGACACACCTCCGCGCGCCAGCTCCCATCCACCGT 335  
Qy 1484 ArgAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db 334 CATCATCCAGTCCCATTCATCTCATCGTCATCGTGTGCT----- 296  
Qy 1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
Db 295 -----CATCATAGTCATATTCCTCAATCTCCACCGCGCGCGGTCCGC 251  
Qy 1524 ArgSerMetLysGlnAlaArgGlnPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 250 CGTCGGCTCTCCAGCACCGCCAG-----CTC 224  
Qy 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGlnArgAsnCysPheGlnLeuLeu 1563  
Db 223 CCCAGGCGCTCCGCGCCCTCCGACACCGTAAGAC-----CTG 185  
Qy 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 184 CACTGCTCCATGCCGA-ACGCAACACCATGCCCATGATCATAAACCATGCCAGGCATA 126  
Qy 1584 GlnAlaLeuMetThrPro 1589  
Db 125 CAGAAAGTCTTAGTCT 108  
  
RESULT 12  
US-08-864-038A-2/c  
; Sequence 2, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; STREET: Isshinden  
; ADDRESS: 812-5 Hirano  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: P-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3331  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA to mRNA



Qy	1119	IleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgIysAsnValIleCysLeu	1133
Db	1397	CTCCGTTACCACCTCCGTTACCACCACCGCTCCACCTCCAGCTGAAGTGAATGCAGATG	1338
Qy	1139	GlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArg---	1157
Db	1337	CAGATG---CTGAAGCAGATGCTGAGCTTTCAATAGACCTGCAGAAGATCCTTTAAGACAT	1281
Qy	1158	-----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMet	1175
Db	1280	CTTTCAAACTGTCATTTTGAGATTTTGCTGCTACTGCTGATGCTTTTAGCAGAGGCTT	1221
Qy	1176	IleGlnAspArgAsnAlaTyrLeuArgThrArgIysCysIleIle-----	1190
Db	1220	TTGCAGAGGCTCCATTACTCTGAGAAGGCTTTAATGTCTCTAATGCGTCCCTTATAC	1161
Qy	1191	-----AsnValGlnArgArgTyrArgAlaThrLeuGlnMet	1202
Db	1160	CATCATTTGTCTAGAACTGAACGATACGACGCGCTGCAGCGGCAGCGCAGC	1101
Qy	1203	ArgArgGluArg-----LysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln	1220
Db	1100	CAGGCTCTCGCTCTCTCCACCACTCCACTCTCCGCCACACCGAGCGGCT---	1044
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLeuArgAlaGlu-----Phe	1236
Db	1043	-----CGCGCGCAGCTGCAGCAGCGGCAGCAGCAGCAGCTCTCGGCGCTC	999
Qy	1237	LeuGlnLeuLysLysValThrLeuValValGlnLysArgArg-----AlaLeuLeu	1254
Db	998	CTCTTCCTCCACCGCGCTACCTCCTCGTCCACCGTAGAAACCAACACCAACCGAGTCCTC	939
Qy	1255	GlnMetArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGln	1274
Db	938	CACCTCCGCAGCAGCGCGCGCAGCAGCAGCGGCAGCAGCAGATCTCCATATCCTC	879
Qy	1275	ArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGln	1294
Db	878	CGA-----GGCCACCAAGACCTCCGAGGCTTCCAGTCGCCACCGAGTCCTCCAA	828
Qy	1295	AlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgIysArgGluArg	1314
Db	827	GTCGCGCCACGAGTCTCTCAAAGTCCACTCGC-----CTCTCGAGCGGCAGCAGCGG	774
Qy	1315	AsnSerPheLeuGlnLeuArgGln-----AlaAlaIleThrLeuGln	1328
Db	773	-----CTCGGCTGCAGCAGAGCTCTCCAACTACTCTGCGCTCTCTCTGCGAG	723
Qy	1329	ArgArgTyrArgAlaArgLeu-----AsnMetIleLysGlnLysSerTyrAlaGlnLeu	1347
Db	722	CGGCAGCAGCGGTGGGTGGGCAGCAGCTCCACCACTCCACCCCGCGGCAGCAG	663
Qy	1348	LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln	1367
Db	662	CGGCAG-----	657
Qy	1368	ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArgGly	1387
Db	656	-----CTGCAGCAGCAGCAGCAGC-----CTCCAGCTCCACCGATCCACCGAGG	609
Qy	1388	AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg	1407
Db	608	CAA-----GAGCTGCACCAAGATCATCGAAT---CTAATCA	573
Qy	1408	LeuGlnLysTyrTrpArgSerIleArg---AspMetArgLeuCysLysAlaGlyTyrArg	1426
Db	572	ATAAATCGTCTCCACGCGCAAGTCGCGCAAGACCTCCAAGTC-----CGC	528
Qy	1427	ArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArg	1446
Db	527	CAAGTCCACCTCCGAGACTC-----CGCCCAATCCAGTCCGAC	489

```

QY 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466
Db 488 CTCAGACACTGCTCTGCTC-----CTGGCTGTCTCCAGACACAGCAC 444
QY 1467 ArgAlaThrLeuLeuMetArgGlnGln-----ArgArgGluPheGluMetLysArg 1483
Db 443 CAGCACCACCTCGCGGACACACACACACACCTCGCGGCGCAGCTCCCATTTCCACCGT 384
QY 1484 ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503
Db 383 CATCATCCCGAGTCCCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 345
QY 1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523
Db 344 -----CATCATAGTCATATTCCTCCACCGCGCGCGCGTCCGC 300
QY 1524 ArgSerMetLysGlnAlaArgGlnGlnPheValGlnLeuArgThrIleAlaValHisLeu 1543
Db 299 CGTCGGCTCTCCAGCACCGCCAG-----CTC 273
QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563
Db 272 CCAGAGCGCTCGCCCTCCGACACCGTAAGAC-----CTG 234
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583
Db 233 CACTGTCTCCATCGGAA-ACGCAACCATGCCCCTGATCATATAACCATGGCAGGCATA 175
QY 1584 GlnAlaLeuMetThrPro 1589
Db 174 CAGAAAAGTCTTAGTCT 157

RESULT 13
US-08-864-038A-4/C
; Sequence 4, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isehinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3331
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: CDS
; LOCATION: from 50 to 2263
; IDENTIFICATION METHOD: P (by similarity to some other pattern)
US-08-864-038A-4

Alignment Scores:
Pred. No.: 2,49e-14 Length: 3331
Score: 257.00 Matches: 167
Percent Similarity: 39.95% Conservative: 147
Best Local Similarity: 21.25% Mismatches: 274
Query Match: 2.70% Indels: 199
DB: 3 Gaps: 35

US-09-914-698-1 (1-1861) x US-08-864-038A-4 (1-3331)
QY 895 LeuHisValValIleGlnArgArgIleArgHisLys-----GluLeuMetArgArgHis 912
Db 2192 ATCATTTGTTTCCACCCGCCCTGATCCGCCACCGCTGAACACACAGATCCGCGGCAC 2133
QY 913 ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys 932
Db 2132 CGC-----CTGATCCACCATCCACCGCACCTCCAGATCCGCGCACCGGATCCAC 2082
QY 933 LeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyr 952
Db 2081 CTCGCGGATCCGCCACACCATCCGACCTCCATTCCTGTTTCCACCGCAATC 2022
QY 953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTyrTrp 972
Db 2021 CATATCGAAATCAGTTCATCAGCACCCTAC-----CGTACATAGCTG 1977
QY 973 ArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAlaAla 992
Db 1976 CGGCTGCAGCAGCGGCGGACACATCAGCAGCATCTCCAGCAGCAGCAGCAGCGG 1917
QY 993 IlePheLeuGln-----ArgIleTrpArgArgArgLeuPheAlaLysLys 1007
Db 1916 CGCGGGCTGCAGCTACAGCACTAGTCTCGTTACGTCGCCATCGCTCCAGTCTCTTAC 1857
QY 1008 LeuLeuAlaAlaAlaGluThrAlaArgGlnArg-----SerGlnLysGlnGlnAla 1025
Db 1856 CTCTACCTCTTCGTCCACCTCCACCAATCCAGCGGCGGCTGCAGCAGCGCAGCAGCTG 1797
QY 1026 AlaAlaSer----- 1028
Db 1796 CTGCTGATGATCCACACCAAAATCTCTCCGAATCTCCACCGAGACTACTCCGAATC 1737
QY 1029 -----TyrIleGlnMetGlnTrpArgThrTyrGln 1038
Db 1736 CTCACCCCATTTCCACCTCCCATCCATCCATCCAGCAGCAGCTGCAGCAGCAGCAGCGG 1677
QY 1039 LeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeu---IleMetPheVal 1057
Db 1676 CAGCAGCGGATCCACCTCCAGCATTTGTCTTCTCAAGGCCCTTCTAAGTGTCTTCTCTC 1617
QY 1058 GlnArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeu 1077
Db 1616 CACCAC-----CTGATGCAGCAGCGGCGGCTGCAGCAGCTGCAGCAGCTG 1578
QY 1078 Lys-----ArgAlaAlaIleAsnIleGlnGlnArg----- 1087
Db 1577 CGGCAGATCCCCACCAAGACCGCTAGTCTCCAAAGTCCCAAAACCTCTCTCCACCTG 1518
QY 1088 -----TyrArgAlaLysLeuSerMetArg-----LysCysAsn 1098
Db 1517 CACCAGCAGCAGTAGTGCAGCGGCTAAGCTCCGCTCCGCTCCACCTCCCAAGTCCAC 1458

```

QY 1099 AlaAspTyrLeuAlaLeuArgSerSerValIleuLysValGlnAlaTyrArgLysAlaThr 1118  
Db 1457 CTCGGCTCTCCGACGACGAGCAGCAGTAGAGCTCCAGCTCCAGCGCGCTCTC 1398  
QY 1119 IleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeu 1138  
Db 1397 CTCGGTTACCACTCCGTTACCAACCGCTCCACCTCCAGCTGAAGCTGATGAGATG 1338  
QY 1139 GlnGlnArgLeuAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArg 1157  
Db 1337 CAGATG---CTGAAGCAGATGCTGAGCTTTTCAATAGACCTGCAAGATCTTTAAGACAT 1281  
QY 1158 -----LeuArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnMet 1175  
Db 1280 CCTCAATCTCAATTGAGATTTTGGCTGCTACTGCTGATGCTTTAGCAGAGGCTT 1221  
QY 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIle----- 1190  
Db 1220 TTGCAGAGCTCCATCTCTGAGAAGGTCTTTAATGTCTCTTAATGCTGCTCCCTTATAC 1161  
QY 1191 -----AsnValGlnArgArgTyrArgAlaThrLeuGlnMet 1202  
Db 1160 CACTCATTTGCTAGAGCTGAAGCAGATGAGCAGCGCTCCAGCGCAGCGCAGCAGC 1101  
QY 1203 ArgArgGluArg-----LysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln 1220  
Db 1100 CAGCGCTCCGCTCTCTCTCCACCACTCCACCTCTCCGCGCACCAACAGCGGTG--- 1044  
QY 1221 IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGlu-----Phe 1236  
Db 1043 -----CGCGCGAGCTGCGAGCGCAGCGCAGCAGCAGCTCTGCGGCTC 999  
QY 1237 LeuGlnLysLysValThrLeuValValGlnLysArgArg-----AlaLeuLeu 1254  
Db 998 CTCTTCTCGACCGGTCTACCTCTCTCTCCACCGCTAGAAACCAACACCGAGTCTCTC 939  
QY 1255 GlnMetArgLysGluArgGlnGluLysLeuHisLeuArgGluValThrIleLysLeuGln 1274  
Db 938 CACTTCCGGCAGCAGCGCGGCGCAGCAGCGCAGCAGCAGATCTCTCATATCTC 879  
QY 1275 ArgArgPheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGln 1294  
Db 878 CGA-----GGCCACCAAGACTCCGAGGCTCCAGTCCGCGCAGCGCTCTCCAA 828  
QY 1295 AlaAlaValSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArg 1314  
Db 827 GTCCGCCACCGAGTCTCTCAAGTCCACCTGCGC-----CTCTGCGCGCAGCAGCGG 774  
QY 1315 AsnSerPheLeuGlnLeuArgGln-----AlaAlaIleThrLeuGln 1328  
Db 773 -----CTGCGGCTGCGAGCAGCTCTCCAGTCTACCTCGGCTCTCTCTGCGAG 723  
QY 1329 ArgArgTyrArgAlaArgLeu-----AsnMetIleLysGlnLeuLysSerTyrAlaGlnLeu 1347  
Db 722 CGGCGAGCGGCTGCGGCTGCGGCGAGCTCCACCACTCCACCCCGCGCGCAGCAG 663  
QY 1348 LysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGln 1367  
Db 662 CGGCAG----- 657  
QY 1368 ValValLeuTyrGlnLysGlnArgGluAlaIleLysValGlnArgArgTyrArgGly 1387  
Db 656 -----CTGCAGCAGCAGCAGCAGC-----CTCCAGCTCCACGCTCCACCGAGG 609  
QY 1388 AsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArg 1407  
Db 608 CAA-----GAGCTGCACCAAGATCATCGAAT---CTAATCAA 573  
QY 1408 LeuGlnLysTyrTyrArgSerIleArg-----AspMetArgLeuCysLysAlaGlyTyrArg 1426  
Db 572 ATAAATCGTCTCCACCGCAAGTCCGCGCAAGACCTCCCAAGTC-----CGC 528

QY 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLysTyrArgAlaThrValGlnAlaArg 1446  
Db 527 CAAGTCCACCTCCGAGACCTC-----CGCCCAATCCAGTCCGA 489  
QY 1447 ArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheIle 1466  
Db 488 GTCCAGCACCTGCTCTGCTC-----CTGCGCTGCTCCAGCAGCAGCAGC 444  
QY 1467 ArgAlaThrLeuLeuMetArgGlnGln-----ArgArgGluPheGluMetLysArg 1483  
Db 443 CAGCACCACTCCGCGCAGCAGCAGCAGCAGCAGCTCCGCGCGCAGCTCCCATTCACCGT 384  
QY 1484 ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db 383 CATCATCCAGTCCCATTCATCATCGTCTCTCT----- 345  
QY 1504 GlnAspTyrGlnLeuIleGlnSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
Db 344 -----CATCATAGTCATATCCCAATCTCCACCGCGCGCGCGCTCGC 300  
QY 1524 ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 299 CGTGGCTCTCCACGACCGCGAG-----CTC 273  
QY 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db 272 CCAGCGGCTCCGCGCTCCGACCGTAAGAC-----CTG 234  
QY 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 233 CACTGTCTCATGCGGAA-ACGCAACCATGCTCATGATCATATAACCATGCGCAGGCATA 175  
QY 1584 GlnAlaLeuMetThrPro 1589  
Db 174 CAGAAAGTCTTAGTCTCT 157

## RESULT 14

US-08-956-171E-63

; Sequence 63, Application US/08956171E

; Patent No. 6593114

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; Gil H. Choi

; Patrick S. Dillon

; Craig A. Rosen

; Steven C. Barash

; Michael R. Fannon

; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 5256

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/956,171E

; FILING DATE: 20-Oct-1997

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/009,861

; FILING DATE: January 5, 1996

; APPLICATION NUMBER: 08/781,986

; FILING DATE: January 3, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Mark J. Hyman

; REGISTRATION NUMBER: 46,789



QY 568 -----GluLeuValValAlaProThrLys 575  
 Db 3605 TGAACAAAGATGGTGGCAACAAACAAAGTAACTCAAGAGTTGATAATATGATCAATTCACAA 3664  
 QY 576 GluGluGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAla 595  
 Db 3665 CAGCATCAATTCAGAAAT-----ACAAATACTTAAACAAC 3700  
 QY 596 ValGluLeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrVal 615  
 Db 3701 AAGCAGACTATTATTAATCAAGATGCAATCATCTTAAACTGCAATCGTGCATCTC 3760  
 QY 616 AsnLysGlnAlaLeuArgIleArgSerAspAsnLeuHisLeuAspValValMetGln 635  
 Db 3761 AAGCGATATTCATGGTTAGTAATAATTACAAAGCTGCATTAATGATTAATCAAGCAG 3820  
 QY 636 ArgThrIleLeuGluLeuLeuCysPheAsnProLeuTyrLeuArgLeuGlyLeuGlu 655  
 Db 3821 CAATTGCTGAATTAGATA-----CTAAAGCTCAAGAA- AAG 3855  
 QY 656 ValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThr 675  
 Db 3856 GTTACAGCAGCACAAACAAAGTAAAGTACGCAAGATGAAGTTGCCAGCACTTGTAACT 3915  
 QY 676 PheIleLeuAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyr 695  
 Db 3916 AAAATTAAACATGATAAAATAATGCAATCCGAGAAATTAATAAAACAACTACACACAA 3975  
 QY 696 ThrLeuThrGluGlu-----TyrAlaGlu 703  
 Db 3976 GGTGTCAACCTGAAAGATGAATGGTATCGCAGTGTAGAACAGATCTGATTACACCA 4035  
 QY 704 ThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAla 723  
 Db 4036 ACAGTTAAACCTCAAGCGCAACAAAGATATATCCAGCAGTT-----ACAACT 4083  
 QY 724 LysGlnLysArgIleValLysHisAsnProCysLeuPheValLysLysSerProHisLys 743  
 Db 4084 CTTAAACAAACAAATTAAGTCAATTCATCATTA-----CAA 4122  
 QY 744 GluThrLysAspIleLeuArgPheSerSerGluLeuLeuAlaAsn-IleGlyAspIle 763  
 Db 4123 GATGAAAGATGTA-----GCAATGATATAAATGGT 4155  
 QY 763 eThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGln 783  
 Db 4156 AAAATTGAACAAAGCGCAATTAAGATATTCATGCGCAACAAACAA----- 4201  
 QY 783 uPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGlyValArgLeuThrAr 803  
 Db 4202 -----ATGCAACAGTAGAAGCCATTAAAAACA 4227  
 QY 803 gValValGluValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProAlaIle 823  
 Db 4228 AAAGCAATCAATGATATTAATCAACTACCTGCTACACAGCTTAAGCAGCAGCTCTT 4287  
 QY 823 eSerArgLeuGlnArgIlePheAsnValLysLeu----- 834  
 Db 4288 GAAGAAATTTGACGAAGTGTCTTCAAGCAACAAATTTGATCAAGCACCTTTAAATCCTGATACA 4347  
 QY 835 -----AlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGln 847  
 Db 4348 ACAATGGAAGATGCGGAAGCTATTGAACCTATTAAATGCAAGTAAAGTTTCTGCTGTT 4407  
 QY 847 yGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLe 867  
 Db 4408 AAAGCAATGGAAGCAGCAGCTG-----CACAG----- 4438  
 QY 867 uTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGln 887  
 Db 4439 -----ATTAGAAAGAGTAAACAAAGAAATCTCAAAATTTGAAAAATATTACT 4488

QY 887 nLysTrpTrpArgArgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGln 907  
 Db 4489 GACTCTACGCAACAAATAATGATGCTTATATGAGTTAAACAAAGCTGCAACACTAGA 4548  
 QY 907 uLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMe 927  
 Db 4549 AAAGCTCAAA-----ATGCTACAG-----TTTCAATGCAACAAAT 4584  
 QY 927 tArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLys 947  
 Db 4585 GAAGAAGTAGCAGAGCTGATGACAG-----TAGATCAGCT 4623  
 QY 947 sPheThrArgArgTyrLeuAlaGlnLysGlnLysTyrGlnSerTyrHisSerIleIleTh 967  
 Db 4624 CAAAGCAAGTTTACATGACATCCAAAGTTGTTAAATCAAAACAGGAAGTCTGCTATACA 4683  
 QY 967 rIleGlnArgTrp---TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheVa 986  
 Db 4684 AAATCAAAAGTATTAGATAAAATCAATGATCAATCAACACAAAGAAAGTTAAACCTGCA 4743  
 QY 986 lGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLys 1006  
 Db 4744 GCTGATACGGAAGTAGAAAACGCAATATATACACGTAAACAAAGAAATTCAAAATAGCAAT 4803  
 QY 1006 sLysLeuLeuAlaAlaAlaGluThrAlaArgLeuGln-----ArgSerGlnLys 1022  
 Db 4804 GCTTCAACTACAGAAAGAAACAGCTGCATATACAGAAATAGATACATAAAAGCAAGAA 4863  
 QY 1022 sGlnGlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIle 1042  
 Db 4864 GCAAGAA-----CAAACTCTGATGCTGCAAAATACAAACAGTGATTAACAACAGCT 4914  
 QY 1042 eGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe-----ValGln 1058  
 Db 4915 AAAGCAATAGTATTGCTGCA-----TTAATCAAGTTCAAGCTGCCCAACAGCT 4965  
 QY 1058 nArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLys 1078  
 Db 4966 AAATCGATGCAAGCGGAATCGCTCAAAAGACAGTGAACGTAAACACAGCAATTTGAA 5025  
 QY 1078 sArgAlaAlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSerMetArgLysCysAs 1098  
 Db 5026 GCAATGAATGATTCGACTACTGAAAGCAACAAACAGCAGCGA----- 5065  
 QY 1098 nAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGlnAlaTyrArgLysAlaTh 1118  
 Db 5066 -----AAAGCAAAAGTGGATCAAG---CAGTAGTTACTGCAACAGCT 5103  
 QY 1118 rIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLe 1138  
 Db 5104 GATATAGATATGCTGACAGCAACAAATGATGTGATATGCAAAACTTACAAATGAAGCT 5163  
 QY 1138 uGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsnTyrLeuArgIle 1158  
 Db 5164 ACAATCGCAGCATTTACACCTGATGCAAAATGTTAAACACAGCAGCAACAAACAAAG- 5215  
 QY 1158 uArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAs 1178  
 Db 5216 -----CAATTGCGATAAAGTACAGCTCAAGCTCAAGAACCAAGTAAATGATGGA 5259  
 QY 1178 pArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTyrArgAl 1198  
 Db 5260 AATAAGCTCTAACACTGAAGAAAHACAGCTGCTAAACCAACAAAGTTCAAACTGAAAAA 5319  
 QY 1198 aThr-----LeuGlnMetArgArg-----Gln 1205  
 Db 5320 ACAACAGCTGATCGCAATAGATGACGACATACATAAATGCGGAAGTTCAAGCGCTAAA 5379  
 QY 1205 uArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGlnIleLysPheArgAl 1225  
 Db 5380 AAAGCAGCAATGCTTAAATTTGAAGCGATTGAGCCAGCAACAAACAACTAAAGATAATGCG 5439  
 QY 1225 aLysArgGluMetLysLysGlnArgAlaGluPheLeuGln-----LeuLysLysValTh 1243

Db 5440 AAAGAGCAATTCCTACGAAGCGAATGACGTAACACACCAATCGCTCAACGCAAGAC 5499  
QY 1243 rLeuValValGlnLysArgArg 1251  
Db 5500 ATTACTGCTGAAGAATTCGAGCGGCTTAATGCGGACGTAGATATGCTGTGACACACGCA 5559  
QY 1252 ---AlaLeuLeuGlnMet 1262  
Db 5560 AATAGCAACATGAGCTGCTAATAGTCAAAATGATAGACCAACGCGAAGACACAG-- 5617  
QY 1262 uTyrLeuHisLeuArgGluValThrLys 1277  
Db 5618 ---GTGAAATAGTATTGATCAAGTACACCAACAGTAAATAAAGCAACT 5667  
QY 1277 eHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaLava 1297  
Db 5668 GCACGTAATGAATCAACAGCAATTT 5700  
QY 1297 lSerCysLeuGlnMetHisTyrArgAsnHisLeuLeuArgLysArgGluArgAsnSerPh 1317  
Db 5701 AAATTGCAAGAGATTCAAGCTACGCCAG-----ATGCAACAGATGAAGAAACACAGCA 5754  
QY 1317 eLeuGlnLeuArgGlnAlaAlaLleThrLeuGlnArgTyrArgAlaLysLysAlaMe 1337  
Db 5755 GCTGATGCTGAAGCAATA-----CTGAAAT 5781  
QY 1337 tLleLysGlnLeuLysSerTyrAlaGln 1346  
Db 5782 GSTAAAGCAATCAAGCCATTTCAGCAGCAACTACTAACCCACAGTGTGATGAAGCTAAA 5841  
QY 1347 ---LeuLysGlnAlaAlaLleThrLeuGlnArgTyrArgAlaLysLysAlaMe 1364  
Db 5842 GCAAATGCAGACAGCGATTATGCGGTAAACCAAAAGTTGTGAAGAAACAGCGGCT 5901  
QY 1364 tGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaLleLleLysValGlnArgAr 1384  
Db 5902 AAAGATGAATTCATCAATTAC---AAGCAACGCAAAACAATGTTATCAATAATGATCAG 5958  
QY 1384 gTyrArgGlyAsnLeuGluMetArgLysGlnLleGluValTyr-----GlnLysGI 1401  
Db 5959 A-----ACGCTACAGAGAAAGAAAGACAGACTTCAACAAATTAGACACAGCA 6009  
QY 1401 n-----ArgGlnAlaValLleArgLeuGlnLysTyrTrpArgSerLleArgaspMe 1418  
Db 6010 GTTACAGACCGGAAATAATATATACAGTGCACCTG-----AT 6048  
QY 1418 tArgLeu-CysLysAlaGlyTyrArgArgLleArgLysSerSerLeuSerLleGlnArgL 1438  
Db 6049 GATAATGGTGTAGATCAGGC-----GAAAGAC 6075  
QY 1438 ystTrpArgAla-ThrValGln-AlaArgArgGlnArgGlu----- 1450  
Db 6076 GCTGGAAGAAATTCATATTCAGACGCAACCCAGCAGCGGTAAATCAAAATGCTAAA 6135  
QY 1451 -lIlePheLeuSerThrLleArgLysValArgLeuMetGlnAlaPheLleArgAlaThrLe 1470  
Db 6136 AATGATGTTGATCAAGCTGCACTCAAAATCAAGCAATTGATTAATACACTGGTCT 6195  
QY 1470 uLeuMetArg-----GlnGlnArgArgGluPheGluMetLysArgArgAlaLava 1487  
Db 6196 ACAACTGAAGAGAAAAATGAGCAAAAGAVTTAGTTTAAAGCTAAAGAAAAAGCGTAT 6255  
QY 1487 lValLleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTyrGI 1507  
Db 6256 CAAGATATCTTAATGCACAAA-----CAACTAATGATGTTACGCAAAATTAAGATCAAGCA 6312  
QY 1507 nLeuLleGlnSerSerValLleLeuValGlnArgLysPheArgAlaAsnArgSerMet-- 1526  
Db 6313 GTTGCTGATATTCAGGTATTACTGCAGATACAACTTAAAGATGTTGCGAAGATGAA 6372  
QY 1527 -----LysGlnAlaArgGlnGluPheValGlnLeuArgThrLleAlaValHisLe 1543

Db 6373 TTAGCAACAAAGCAAAACGAAACAAAGACGCTTAA---TTGCACAAATCGCAGATGCGACT 6429  
QY 1543 uGlnGlnLysPheArgGlyLysArgLeuMetLleGluGlnArgAsnCysPheGlnLeuLe 1563  
Db 6430 ACTGAAGAAA---AAGAACAGCAAAATCAACAAAGTAGACCCCAATTAACACCAAGGTAAT 6486  
QY 1563 uArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPh 1583  
Db 6487 C-----AAAAT 6492  
QY 1583 eGlnAlaLeuMetThrProGluMetMetAspLeuLleArgGlnLysArgAlaLysVa 1603  
Db 6493 ATTGAAAATGCACAGTCAATCGATGATTAACACTGCAAAAGATATGCAATTCACGCA 6552  
QY 1603 lLleGlnArgTyrTrpArgGlyTyrLeuLleArgArgGlnLysHisGlnGlyLeuLe 1623  
Db 6553 ATTGACCCCAATTC-----AAGCATCAACAGATGTT 6582  
QY 1623 uAspLleArgLysArgLleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValAr 1643  
Db 6583 A-----AACGAATG-----CAAGAGCGGAATGC-----TAACGTAA 6615  
QY 1643 gCysLysValGlnGlnGluAlaValArgPheLeuArgGlyArgPheLleAlaSerAspAlaLe 1663  
Db 6616 ATGCAAAATAAAATAACTGAAATACTTAATAATAATGACACTACTAATGAAGAAAAAGGT 6675  
QY 1663 uAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCy 1683  
Db 6676 AAGCATTTGGACAGTTAGACGAGCATATGAAGAGGTTTAAATTAATATTATG----- 6730  
QY 1683 sSerGluPheMetSerThrPheCysTyrGlyLleMetAlaGlnAlaLleArgSerGluVa 1703  
Db 6730 ----- 6730  
QY 1703 lAspLysGlnLeuLleGluArgCysSerArgLleLleLeuAsnLeuAlaArgTyrAsnSe 1723  
Db 6731 ---CAGCAACTACTACAG---GTGATGTAACACTACTGCTAAAGATACAGCAGTACAAAA 6783  
QY 1723 rThrThrValAsnThrPheGlnGluGlyLeuValThrLleAlaGlnMetLeuLeuAr 1743  
Db 6784 G-----TTCAACAACATTCATG-----CAAACTCTGTTAAG 6813  
QY 1743 gTrpCysAspLysAspSerGluLlePheAsnThrLeuCysThrLeuLleTrpValPheAl 1763  
Db 6814 A-----AACGACAGCTAAAAAGAAATAGATCAAGCTGCAGCTGATAAGAAA 6861  
QY 1763 aHisCysProLysLysArgLysLleLleHisAspTyrMetThrAsnProGluAlaLleTy 1783  
Db 6862 ACACAAATAGAACAAACACCAATGCAATCACAACAGAAA----- 6901  
QY 1783 rMetValArgGluThrLysLysLeuValAlaAlaArgLysGluLysMetLysGln 1800  
Db 6902 -TTAATGATGCAAAACACAGAGTTGATGACTGAATTAATCAACGGAACAA 6952

RESULT 15  
US-09-976-594-907  
; Sequence 907, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 907  
; LENGTH: 5923  
; TYPE: DNA  
; ORGANISM: Homo sapiens

```

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 977975.1
US-09-976-594-907

```

Alignment Scores:	
Pred. No.:	9,749-14
Score:	25.50
Percent Similarity:	35.13%
Best Local Similarity:	19.77%
Query Match:	2.69%
DB:	4
Length:	5323
Matches:	291
Conservative:	227
Mismatches:	531
Indels:	426
Gaps:	61

US-09-914-698-1 (1-1861) x US-09-976-594-907 (1-5923)

252 ThrHisThrArgAlaLeuAlaCysIleHisGluGluGlyProSerProProArgThr 271  
941 ACAATTCAGAGTGAAGTAATGGAATGCATCAGAGGAGCATCA----- 988  
272 ProThrLysSerAlaIleHisAspLeuLys-ArgAspIleLysLeuValGlySerProLe 291  
989 ---AAATATTCAGGATTGGAGATTGAAATGAAGATTGGAAA-----GAGAGGTT 1036  
291 uArgLysTySerGluSerMetLysAspLeuSerLeuSerProGlnThrLysTyAl 311  
1037 GAGAAAAATTCAGCAAGAACAAAGAACTT-----TTGGATAA 1075  
311 alleGlnGlySerMetProAsnLeuLenglu-----MetLysIleArgSerIleGluGl 329  
1076 AGTCATGTTTACAGTTACAGCTGATGAGGAAGTTATGGTGTCTGATGATCTGGAAG 1135  
329 nAsnArgTyTyGlnGluGlnGlnIleLysAlaLysAspLeuAsnSerSerSe 349  
1136 CGAGAGA-----GAAAAGCTGAAGTCCCTTTTGGCAGCTTAAGAAAGCAACATGAAGA 1189  
349 rSerSerGluAlaSerLeuAlaGlyGlnGlnGluPheLeuPheAsnHisSerGluIleLe 369  
1190 AAGCTTAAAGACTATTGAGGCTCTGAAAATAAGTAGATTAAATATTTTCAGAGTGCATATT 1249  
369 uAlaGlnSerSerArgPheAsn-----LeuHisGluValGlyArgLys 383  
1250 AGATCAGAGAGTCATTTCAGTACACCAAGAAAGAGATATGCTTTCTTAAACAGGTCAGAT 1309  
383 sSerValLysGlySerProValLysAsnPro-----HisLysArg----- 396  
1310 GTATATGGCAGACTCAGCTGTACTTCCCGAGTATACAGCCCATATGCAAGCAGATC 1369  
397 -----ArgSerHisGluLeuSerPheSerAspAlaProSerAsnGluSerLeuTyAr 414  
1370 TATGTTTAAGACCTCTCGAAGTATCTTTA-----CCAGTCAACAGCTCATCTCTGA 1420  
414 GAsnGluThrValAlaIleSerProProLysLysGlnArgValGluAspThrThrLeuPr 434  
1421 AAATGAAATTTTA-----AAGAAGAGTTAGAAGCAATGCGAATCTTCTG 1465  
434 oArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerSerAlaHisAlaTrpProHi 454  
1466 TGAGTTCAGCAAAACAAGACCGACTGAAGCTCCAAAATGAATGGCACAC----- 1514  
454 sAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaTh 474  
1515 -----AAAGTGCAGAAATGCAAAAGCTTTAGCATTTAGCATTTAGAAAG 1555  
474 rProArgLysValArgAspThrSerIleGlnPro-----Se 486  
1556 GGTCAGGAGGATTACAGATCAACAGATAAGCAATTAGAAGATGCATTTAAAGATGTGCA 1615  
486 rValLysLeuTyAspSerGlu-----LeuTyMetGlnThrCysIleAsnProAs 503  
1616 GAAGAGGATGTATGACTCAGAGGTAAAGTTAAACAAATGCAGACC----- 1661  
503 pProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyLeuAspGl 523

1662	Db	-----CATTTCTTGCCCTTAAAGAACACTTAACAAG	1693
523	Qy	uGlnAlaValAspArgHisGlnAlaAspPheLysIysTrpLeuAsnAlaLeuValSerI	543
1694	Db	...   ...:::	
543	Qy	eProAlaAspLeuAspAlaAspLeuAsnAsnLys-----lLeAspValGl	558
1730	Db	...   ...	
558	Qy	YlysLeuPheAsnGluValArgAsnLysGluLeuValAlaProThrIysGluGluGl	578
1790	Db	...   ...	
578	Qy	nSerMetAsnTyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaValGluLe	598
1850	Db	...   ...	
598	Qy	uPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsnLysGl	618
1892	Db	...   ...	
618	Qy	nAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrI	638
1952	Db	...   ...	
638	Qy	eLeuGluLeuLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuGluValValPh	658
1971	Db	...   ...	
658	Qy	eGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLe	678
1994	Db	...   ...	
678	Qy	uAsnArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuTh	698
2022	Db	...   ...	
698	Qy	rGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuPr	718
2060	Db	...   ...	
718	Qy	oPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheVally	738
2114	Db	...   ...	
738	Qy	sLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAl	758
2156	Db	...   ...	
758	Qy	aAsnIleGlyAspIleThrArgGluLeuArgLeuGlyTyrValLeuGlnHisArgGl	778
2216	Db	...   ...	
778	Qy	nThrPheLeuAspGlu-----PheAspTyrAlaPheAsnAsnLeuAlaValAspLeuAr	796
2270	Db	...   ...	
796	Qy	gAspGlyValArgLeuThrArgValValGluValIle-----LeuLeuAr	811
2330	Db	...   ...	
811	Qy	gAspAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAs	831
2390	Db	...   ...	
831	Qy	nValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAl	851
2438	Db	...   ...	
851	Qy	aAlaGlnAspIleVal-----AspGlyHisArgGluLysThrLeuSerLeuLe	867
2498	Db	...   ...	
867	Qy	cGCGCTTGAACCTGTGTTGTACCTCCCTGAGAAACATGAAAGAGATAATAGCTCTGAA	2557



QY 867 uTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuG1 887  
Db 2558 ATCCAATATTGTT-----GAACTTAAGAAACACGCTGCTCACTTAA 2599  
QY 887 nLys-----TrrTrpArgArgHisTrpLeuHisValValIleGlnArgArgIleArgHi 905  
Db 2600 GAAAAAATGGTGGTGAACACGAGAGAAATACACGCTCTCACATCTGAAAAACACTAACTT 2659  
QY 905 sLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHi 925  
Db 2660 GAAGAAGATGATG----- 2672  
QY 925 sGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIle 943  
Db 2673 -----AGTAATCAGTATGTCGCGCTAAACCCATGAAGAGGTAAATGACACTGAATGA 2728  
QY 944 -IleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHi 963  
Db 2729 CACGTTAGCCCAAACTAACAGAGAAATTATTAGATGTGAAGAAAAAATTGAAGATATAAA 2788  
QY 963 sSerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgG1 983  
Db 2789 T----- 2791  
QY 983 nArgPheValGlnLeuArgGlnAlaAlaIlePheLeuGlnArgIleTrpArgArgLe 1003  
Db 2792 GGAATTTGTAATAAATAAAGATAAGATAAGATGAATATTAAAGA----- 2834  
QY 1003 uPheAlaLysLysLeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysG1 1023  
Db 2835 -----AACCTGAAAAACACTCAG---AA 2854  
QY 1023 nGlnAlaAlaAlaSerTyrIleGlnMetGlnTrpArgThrTyrGlnLeuGlyArgIleG1 1043  
Db 2855 CCAAATAAAGCTGAGTACATCAGC-----CTGGC 2884  
QY 1043 nArgHisGlnPheLeuArgGlnArgAspLeuIleMetPheValGlnArgArgMetArgSe 1063  
Db 2885 AGACACGAG-----GC 2896  
QY 1063 rLysTrpSerMetLeuGluGlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAs 1083  
Db 2897 AAAGATGAGCTCGCTAAGTCAG----- 2918  
QY 1083 nIleGlnArgTrpArgAlaLysLeuSerMetArgLys-----CysAsnAlaAs 1100  
Db 2919 -----AGCATGAGAAAGGTGCAGGATAGTAATGCTGA 2950  
QY 1100 pTyrIleAlaLeuArgSerSerValLeuLysValGlnAlaTyrArgLysAlaThrIleG1 1120  
Db 2951 AATCTTGCC-----AACTACAGAAAGGC----- 2975  
QY 1120 nMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAsnValIleCysLeuGlnG1 1140  
Db 2976 -----CAAGAAGAGATTGTGCACACTGCATGC 3001  
QY 1140 nArgLeuArgAlaIleMetLysMetArgGlu-GlnArgGluAsnTyrLeuArg-----L 1158  
Db 3002 CGAAATTAAGCCCAAGAGAGAGCTGCACACATAAAGAAATGCATTAAAGGTAATAA 3061  
QY 1158 euArgAsnAlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnA 1178  
Db 3062 TGCCCCAATTGTCAGCTTTGAGGATGCGAGAGAAATTAAGAACACAGAGAAAGAACT 3121  
QY 1178 sPArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgA 1198  
Db 3122 AAAAGACCACTTATCAGACGACACACAAAAGTATAGTGTGCTGAGAGAGAGTCAAGAA 3181  
QY 1198 laThr---LeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrL 1217  
Db 3182 AAACAAGCAAGAGATGACAGTTAAAGAGAGAGATTTTACC-CTTCAGAAAGATTGA 3240

QY 1217 ysArgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheL 1237  
Db 3241 GAGATAAGACAGTTCCTCAATTGAGAGCTCTCAATAATGGAAGA----- 3284  
QY 1237 euGlnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetA 1257  
Db 3285 --GCATTAAAGCAGAAAACACAGCAGCTAAACAAA-----CAGTTAA 3324  
QY 1257 rGlysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgP 1277  
Db 3325 AAGACTTGTGCAGAAATACACGGAAGTAAAGAAATGTGAAGAGAGAGCTAGTAGAGAA- 3383  
QY 1277 heHisAlaGlnLysSer-MetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAla 1296  
Db 3384 --AATGCCAAACAGACTTCTGAGATATCTGCAGTGCAAA----- 3420  
QY 1297 ValSerCysLeu-----GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArg 1314  
Db 3421 ATCTTTTGCAAAAACAACTGTTCCATTTGGAACAGGTGAGGCTCTG-AAAAATCTCTT 3479  
QY 1315 AsnSerPheLeuGln---LeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAla 1333  
Db 3480 AATGCCCAATTTGAAATCTAAAGGAAGAACTGAAGAGTATGCAAGGTGTTCAGAGAA 3539  
QY 1334 ArgLeuAsnMetIleLysGlnLeuLysSerTyrAlaGln---LeuLysGlnAlaAlaIle 1352  
Db 3540 GAGCAGCAGACAGTGCACAACTGCATCAATTTGTGAGAAATCAAAAGAACTCTCTCTGTA 3599  
QY 1353 ThrIleGlnThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGln 1372  
Db 3600 CCCCTGCGCAGACGNTTGCAGATTAAAGAGCACTTGCAGAAAGAGTT----- 3647  
QY 1373 LysGlnArgGluAlaIleIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArg 1392  
Db 3648 -----GGAATCATATAAGCCAGCTTGAGAGAAAGAGAGAAAGCAAAAC 3695  
QY 1393 LysGlnIleGluVal-----TyrGlnLysGlnArgGlnAlaValIle 1406  
Db 3696 AAAATGGAAGAAGTCTCCAACTTCAGTCGAGGTTGAGAAATCAATAACAAAGCACTTAA 3755  
QY 1407 ArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArg 1426  
Db 3756 AAATTAGAGACT-----AGAGAGTGTGCTTGTCTAAATATAAGCAACAAAGT 3809  
QY 1427 ArgIleArgLeuSerSerLeuSerIleGlnArgLys-TrpArgAlaThrValGlnAlaAr 1446  
Db 3810 GATTTCGAGACACAGATTTCTAGCTTAAATGAAATAATGGCCAATCTCAATAGAAAGT-- 3867  
QY 1446 gArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGlnAlaPheI1 1466  
Db 3867 ----- 3867  
QY 1466 eArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArgArgAlaAl 1486  
Db 3868 -----ATGAGAAAGTATGTCAGGAAGTTTTCATGCCAAAAAGA----- 3906  
QY 1486 aValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArgGlnAspTy 1506  
Db 3907 -----AGAAATATCTGCAAAAAGATGAGAAAGGAATTA 3938  
QY 1506 rGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsnArgSerMe 1526  
Db 3939 CTGCATTTTCAGCATTCAGCAAGAAATTAAG-GATCAGAAAGGACGATGTGATAGTCTCT 3997  
QY 1526 tLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeuGlnGlnly 1546  
Db 3998 A-----ACACAATCACAGAGTTACAAAGAAG 4024  
QY 1546 sPheArgGlyLysArgLeuMetIleGluGlnArg-----As 1558  
Db 4025 AATACAGAATCTGTAAACAAATAGAAAGCAAAAGATAATAGATACTGAACCTGCTTAA 4084  
QY 1558 nCysPheGlnLeuLeuArgCysSerMetProGlyPheGln-----AlaArgAlaArgG1 1576



```

Db      4085  TGAATGGAAGATTAAACAGGCACTCAATGGCCCTTCCCAACTCACCCTACACAGTGG 4144
Qy      1576  yPheMetAlaArgLysArgPheGlnAlaLeuMetThr-----
Db      4145  GAAACCCACCAAGAGGAGCGAGCGCTGATTGACACTCTGCAGCACCAAGTGAATCTCT 4204
Qy      1589  -----ProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysValIleGlnArgTy 1607
Db      4205  GGAGCAACAGCTGGCCGATGCTGACAGACAGCACCAGAAAGTAATTGCAATT----- 4256
Qy      1607  rTTPArgGlyTyrLeuIleArgArgArgGlnLysHis-----GlnGlyLeuLeuAspI1 1625
Db      4257  -TATCGGACACACCTTCTTAGTGTGTCACAGGGTCAATGGATGAAGATGTTTCAGGAGGC 4315
Qy      1625  eArgLysArgIleAlaGlnLeuArgGln 1634
Db      4316  TCTGCTCCAGATCATACAAATGCGCAG 4343

```

Search completed: September 15, 2004, 13:05:22  
Job time : 1990 secs

*This Page Blank (usp10)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: September 15, 2004, 06:57:11; Search time 1833 Seconds

(without alignments)  
5108.382 Million cell updates/sec

Title: US-09-914-698-1

Perfect score: 9514

Sequence: 1 MELVNSPVLEVACKETLQLI.....FISSVYAFDTILCKLQIDMF 1861

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh

-Q=/cgn2\_1/USPIO\_spoolh/US09914698/runat\_10092004\_150008\_28369/app\_query.fasta\_1.2055

-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09914698 -CGN\_1\_1260 @runat\_10092004\_150008\_28369

-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match Length	ID	Description
------------	-------	--------------	----	-------------

1	834	8.8	2453	13	US-10-342-887-227	Sequence 227, App
2	834	8.8	2453	13	US-10-172-118-227	Sequence 227, App
3	707.5	7.4	3887	13	US-10-342-887-1746	Sequence 1746, App
4	707.5	7.4	3887	13	US-10-172-118-1746	Sequence 1746, App
5	700.5	7.4	4027	17	US-10-115-635-64	Sequence 64, Appl
6	678	7.1	2811	16	US-10-108-260A-956	Sequence 956, App
7	437	4.6	2414	13	US-10-342-887-1748	Sequence 1748, App
8	437	4.6	2414	13	US-10-172-118-1748	Sequence 1748, App
9	437	4.6	2414	17	US-10-188-832-81	Sequence 81, Appl
c	348.5	3.7	3069	17	US-10-437-963-53524	Sequence 53524, A
10	348.5	3.7	3069	17	US-09-814-353-21993	Sequence 21993, A
11	322	3.4	7677	10	US-09-954-456-1601	Sequence 1601, App
12	322	3.4	14800	9	US-09-954-456-1601	Sequence 1601, App
13	322	3.4	14800	15	US-10-269-909-61	Sequence 61, Appl
14	322	3.4	14800	17	US-10-717-597-183	Sequence 183, App
15	317.5	3.3	15231	9	US-09-917-800A-1505	Sequence 1505, App
16	317.5	3.3	15231	12	US-10-152-319A-1780	Sequence 1780, App
17	316.5	3.3	14835	15	US-10-240-965-113	Sequence 113, App
18	316	3.3	6401	15	US-10-171-581-352	Sequence 352, App
19	316	3.3	6401	15	US-10-177-293-314	Sequence 314, App
20	293	3.1	5229	16	US-10-012-697-1457	Sequence 1457, App
21	283.5	3.0	6946	10	US-09-814-353-20816	Sequence 20816, A
22	283	3.0	6839	13	US-10-342-887-1803	Sequence 1803, App
23	283	3.0	6839	13	US-10-172-118-1803	Sequence 1803, App
24	276.5	2.9	5457	13	US-10-282-122A-28061	Sequence 28061, A
25	275	2.9	5895	16	US-10-389-493-26966	Sequence 26966, A
26	273.5	2.9	5418	13	US-10-282-122A-27329	Sequence 27329, A
27	273.5	2.9	8503	15	US-10-037-270-130	Sequence 130, App
28	273.5	2.9	8503	16	US-10-117-722-130	Sequence 130, App
29	273.5	2.9	580073	15	US-10-205-220-1	Sequence 1, Appli
30	273	2.9	7434	9	US-09-815-242-4761	Sequence 4761, App
31	273	2.9	7437	9	US-09-815-242-8869	Sequence 8869, App
32	273	2.9	8493	14	US-10-071-766-51	Sequence 51, Appl
33	271.5	2.9	6075	16	US-10-388-934-532	Sequence 532, App
34	271.5	2.9	6075	17	US-10-664-705-62	Sequence 62, Appl
35	271.5	2.9	7992	10	US-09-893-519A-140	Sequence 140, App
36	271.5	2.9	8491	15	US-10-133-013-260	Sequence 260, App
37	270	2.8	11677	15	US-10-082-830-134	Sequence 134, App
38	268	2.8	10190	9	US-09-864-864-232	Sequence 232, App
39	268	2.8	10211	9	US-09-954-456-1153	Sequence 1153, App
40	268	2.8	10211	9	US-09-967-768A-186	Sequence 186, App
41	268	2.8	10211	10	US-09-918-624B-2	Sequence 2, Appli
42	265	2.8	10281	15	US-10-084-817-361	Sequence 361, App
43	265	2.8	10281	15	US-10-240-965-201	Sequence 201, App
44	260.5	2.7	11736	16	US-10-291-265-218	Sequence 218, App
45	260	2.7	6452	15	US-10-205-823-418	Sequence 418, App

## ALIGNMENTS

RESULT 1  
US-10-342-887-227  
; Sequence 227, Application US/10342887  
; Publication No. US20040058340A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter S.  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Christopher J.  
; APPLICANT: Van 't Veer, Laura Johanna  
; APPLICANT: Van de Vijver, Marc J.  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14



Db 1910 ----- 1910

QY 995 uGlnArgIleTTPArgArgArgLeuPheAlaLysLysLeuLeuAlaAlaGluThrAl 1015

Db 1911 -----CGAAGAGTCTTAGCAGAGAGAAATATTAAATGTTAAAGAGGA 1954

QY 1015 aArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrIleGlnMetGlnTrpAr 1035

Db 1955 -----AAAGCTGGAAAAGTTCAAAATAA-----CGAGATCATTATTCAGGATATGGAG 2008

QY 1035 gThrTyrGlnLeuGlyArgIleGlnArgHISgluPheLeuArgGlnArgAspLeuIleMe 1055

Db 2009 AAGATATTCC-----ACTAGACAAAGATTTCTGAAATGGAATATTATTCAAT 2056

QY 1055 tPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGluPheGl 1075

Db 2057 CATCTGCAATCTAGATAAGA-----ATGATAATTCGTGTACATCTTATAA 2104

QY 1075 nGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetAr 1095

Db 2105 ACGATATCTTTGGGTACAGTTACATTCAGAGGCAATGGCGTCTTATTAAAGAGAA 2164

QY 1095 gLysCysAsnAlaAspTyrLeuAlaLeuArgSerValLeuLysValGlnAla---Ty 1114

Db 2165 ACAAGATCAACAAGATATGAAATGCTAAATCATCAACTCTTATAATCCAACTATGTT 2224

QY 1114 rArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyrSerLeuArgLysAs 1134

Db 2225 CAGAAAA-----TGAAGCAACCGTAAAA-- 2246

QY 1134 nValIleCysLeuGlnGlnArgLeuArgAla---IleMetLysMetArgGluGlnArgGl 1153

Db 2247 -----ATGCATCACAAGTAAAGCTACAGTANATTCGAAAGCTTTAGAGA 2296

QY 1153 uAsnTyrLeuArgLeuArg-----AsnAlaSerIleLeuValGlnLysArgTy 1169

Db 2297 ATGGCATTTAAGAAAACAAGCTAAAGAGAAATTCCTGCTATTATCATCAATCATGTA 2356

QY 1169 rArgMetArgGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIl 1189

Db 2357 TAGAATGCATAAAGAAATTA-----CGGAAGTATATATTTATATAGATCTTGTTG 2404

QY 1189 eIleAsnValGlnArgArgTrpArg 1197

Db 2405 TGTATCATTCAGAAAGATTTCGG 2429

RESULT 2

US-10-172-118-227

; Sequence 227, Application US/10172118

; Publication No. US20030224374A1

; GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue

; APPLICANT: He, Yudong

; APPLICANT: Linsley, Peter

; APPLICANT: Mao, Mao

; APPLICANT: Roberts, Chris

; APPLICANT: Van 't Veer, Laura

; APPLICANT: Van de Vijver, Marc

; APPLICANT: Bernards, Rene

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

; FILE REFERENCE: 9301-175-999

; CURRENT APPLICATION NUMBER: US/10/172,118

; CURRENT FILING DATE: 2002-05-14

; PRIOR APPLICATION NUMBER: 60/380, 770

; PRIOR FILING DATE: 2002-05-14

; NUMBER OF SEQ ID NOS: 2699

; SEQ ID NO 227

; LENGTH: 2453

; TYPE: DNA

; ORGANISM: Homo sapiens

; PUBLICATION INFORMATION:

; DATABASE ACCESSION NUMBER: AK001380

; DATABASE ENTRY DATE: 2001-06-18

US-10-172-118-227

Alignment Scores:

Pred. No.: 3,21e-72 Length: 2453

Score: 834.00 Matches: 235

Percent Similarity: 44.65% Conservative: 153

Best Local Similarity: 27.04% Mismatches: 228

Query Match: 8.77% Indels: 254

DB: 13 Gaps: 22

US-09-914-698-1 (1-1861) x US-10-172-118-227 (1-2453)

QY 522 AspGluGlnAlaValAspArgHISgluAlaAspPheLysLysTrpLeuAsnAlaLeuVal 541

Db 6 GATGAACCTCGAAGGAGAAACAGGAGCCTTCACTTGGTGGTTAAATTTATATTA 65

QY 542 SerIleProAlaAspLeuAspAlaAspLeuAsn---AsnLysIleAspValGlyLysLeu 560

Db 66 ACC---CCTGATGACTTCACTGTAAACAAATATTCTGAAGTAAATGCTGCTACTCTT 122

QY 561 PheAsnGluValArgAsnLysGluLeuValVal-----AlaProThrLysGluGlu 577

Db 123 CTTTGGGAATAGAGAAATCAACATAAAATAAGTGTTCCTAGAGCACCTACAAAAGAGAA 182

QY 578 GlnSerMetAsn---TyrLeuThrLysTyrArgLeuGluThrLeuArgLysAlaAlaVal 596

Db 183 ATGCTCTCAGAGCTTATACCTGCTGGGTAGTTAAACAGACTACGTCGTCAGCAGTC 242

QY 597 GluLeuPheSerGluGlnMetArgLeuProCysSerLysValAlaValTyrValAsn 616

Db 243 CGTTTGTATTCTTGAAGAAATGGTTAAAGCTTAAAGCTTGAATTCGAATTTGAA 302

QY 617 LysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArg 636

Db 303 GCTAGGCGGTAAATGTTGCGAAAGATAGACACCTATCGAAAGATGTGGAGAACGTCAG 362

QY 637 ThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeuArgLeuGlyLeuVal 656

Db 363 AAAGTCCCGAANTGGCTGTTGCTCTACAACTCTTGTGGCTTCGAATTTGGTCTAGAGA 422

QY 657 ValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPhe 676

Db 423 ACTTATGGAGAACTCATATCTTGGAGATTAACAGTATGTCACAGGTTGGCTATGTTT 482

QY 677 IleLeuAsnArgLeuPheArgAsn-----LysCysGluGluGlnArgTyrSerLys 693

Db 483 ATTCTGAATCGCTACTTGTGAATCCTGATATAGCAGCTGAGTATAGACACCCACTGTT 542

QY 694 AlaTyrThrLeuThrGluGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIle 713

Db 543 CCTCAGCTGATAGAGATGGTCATGAAGAGCTTTGTCCAAAGTTTACATTTGAAAGTTA 602

QY 714 LeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnPro 733

Db 603 TTGTTTGGTCTGCTTTCTTGATTAAGCTAAATTTCCAGACTCATTTGATCATGATCCT 662

QY 734 CysLeuPheValLysLysSerProHisLysGluThrLysAspIleLeuLeuArgPheSer 753

Db 663 TGTCTCTCTGTAAGATGCGCAATTCAGGCTAGTAAAGAAATCCTTTAGGCTTTTCA 722

QY 754 SerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrVal 773

Db 723 CGAATTTCCTAAGTGGTGAAGGTGACCTTTCCCGCTCGCTTATTTGGGATTTACCT 782

QY 774 LeuGlnHisArgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaVal 793

Db 783 GTTAACCATGTTCCAGACACCATTTGATTTGATTTGCGGTTTACAAATCTTCCGCTA 842

QY 794 AspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLeuArgAsp 813

Db 843 GACTTGCAATGTGGAGTGGCGCTTGTGCGAACCATGGAATCTTCTCACACAGAACTGGGAC 902



Alignment Scores:

Pred. No.:	2.89e-59	Length:	3887
Score:	707.50	Matches:	276
Percent Similarity:	40.70%	Conservative:	225
Best Local Similarity:	22.42%	Mismatches:	412
Query Match:	7.44%	Indels:	319
DB:	13	Gaps:	40

US-09-914-698-1 (1-1861) X US-10-342-887-1746 (1-3887)

QY	868	TrpGlnLeuLeuTyrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGln	887
Db	359	TGGATGATTAAGAAAGGATGCGA-----GAGATGCACAGGGCTGCTACTTTCATCCAG	412
QY	888	LysTrpTrpArgArgHisTrpLeuHis-----Val	897
Db	413	TCCTACTTTCAGAAATGCACAGATTACATATGAGATATCAGGCTTTGAAAACAGGCCCTCCGTT	472
QY	898	ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet	909
Db	473	GTGATCCAAACAGCAATATACCAAGCAAAATAGAGCTGCAAAATGCAGAGCGCAGCATTTATCTC	532
QY	910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929
Db	533	AGACAAGACACTCTGCTGTGCATCCCTCAGGCTCATTCAGGGTATGAAAACCTGGAAGA	592
QY	930	TyrValLysLeuPheLysThrsGluArgThrGlnAlaAlaIleLeuGlnLysPheThr	949
Db	593	CATTTCAGAGAGTATGCAATTC-----TCTGCAACCCCTATTTCAGAGTAGGTTT	640
QY	950	ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969
Db	641	AGATCATTTACTGGTCAGGAGAAGATTCAATTCCTCCCTCAAAAAGCTACTATTTTTGTTCAG	700
QY	970	ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg	989
Db	701	AGGAAATATCGAGGCCACCAATTGTGTGCACAAACATATAAATTGTACCAATCTTGCACCTTAAGA	760
QY	990	GluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeu---	1008
Db	761	AGGCGCCCATTAACATACAGTCACTTCACAGAAGACTGATGTAAAGAGAAGATTACAA	820
QY	1009	-----LeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln---	1023
Db	821	GAAATGCAAGGGCTGCAGT--TCTCAATCAGGCTACTTTCAGGATGCACAGAAAAA	879
QY	1024	-----GlnAlaAlaSerTyrIleGlnMetGlnTrpArgThr	1036
Db	880	TATATTACATTTACAGCTGGAACATGCTTCAATCTTAATTCAGCAACATTATCGAACA	939
QY	1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe	1056
Db	940	TATAGAGCTGCAAAATTTGCAAGAGAAAAATTATATCAGACAATGGCATTCCTGCTGTGCTT	999
QY	1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGlu	1073
Db	1000	ATTGAGCTGCATATAAGGANTGAAGACAGACAACCTTTTAAGGAAAAACACAAAGCT	1059
QY	1074	-----PheGlnGlnLeu	1077
Db	1060	TCTATTGTAATACAAGGCACCTCAGAAATGTATAGCGAGTATTGTTTCTACCAAAAGCTT	1119
QY	1078	LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097
Db	1120	CAGTGGGCTACAAAATCATCAGAAAAAATPATAGACAAATAAAAAAGAACAGAAA---	1176
QY	1098	AsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----ValGlnAla	1113
Db	1177	-----GTATTTCACACAACTGAACCTTGAAGAAGAGACTTGTGTTCAGGCA	1221
QY	1114	---TyrArgLysAlaThrIleGlnMetArgIleAsp	1124
Db	1222	GGTTTTCAGGACATGAACATATAAAAAACAGATTCAGGAACAGCACCGCTGCCATTATT	1281

Qy	1125	----	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1133
Db	1282	ATTTCAGAAAGCATTTGTATAAGCCTTTAAAATTAAGAAGCATTTATCCACATTAGAGCAACA	1341	
Qy	1135	VaIlleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	1151	
Db	1342	GTAGTTTCTATTCAAAGAAGATACAGAAACTAATCGACGTGGCTGCCAACGCAGTTATT	1401	
Qy	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160	
Db	1402	TGTATACAGTCTTTATTACAGAGCGTTTAAAGTACGAAGAGATATTCAAAATATATGCACCGG	1461	
Qy	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180	
Db	1462	GCTGCCACATAATTCAGTCAATTCATCGAATGCACAGGGCCAAAGTTGAT-----	1512	
Qy	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200	
Db	1513	----TATGAAACAAAGAAACTCGCAATGTGGTTATACAGAAATTATTATAGTGTTGATGTT	1569	
Qy	1201	GlnMetArgGluArgLysAsnTyrLeuHiIleGlnIlnThrThrLysArgIleGln	1220	
Db	1570	AGAGTAAAAACAGAAAGAAAAACTTTTTAGCAGTTTCAGAAATCTGTACCAGTAATTCAG	1629	
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys	1240	
Db	1630	GCTGCTTTTAGAGGCATGAAA-----GTTAGACAAAAATTTGAAATGATCACAGGAA	1683	
Qy	1241	LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260	
Db	1684	AAGATGCACGCCATTGTT-----AACCAATCTGCACTCTGCTGTTTACAGAAAGTAAA	1734	
Qy	1261	GlnGlnTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln	1280	
Db	1735	ACTCAGTATCGAAGCTGTTCAAAGTGAAGTGTTATGATTCAGAGTGGTATFAAGCTTCT	1794	
Qy	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaIleValSerCysLeu	1300	
Db	1795	GGCCTTCTGTTTTCACAGGAAGCAGAGTATCAITCTCAAAGTAGGCGTGCAGTAAACAATT	1854	
Qy	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313	
Db	1855	CAAAAGCTTTTGTAGATGGTTCACAAAGAACTGGAAACACACAGAAATGTGTCGCCTA	1914	
Qy	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322	
Db	1915	CGGATTCAGTTCTTCCTTCAGATGGCTGTGTATCGGAGAAGATTGTGTCAGCAGAAAAGA	1974	
Qy	1323	AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetLysGlnLeuLys	1342	
Db	1975	GCTGCTATCACTTTACAGCATTAATTTTAGACG-----TGGCAAAACCCAGAAA	2022	
Qy	1343	SerTyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362	
Db	2023	CAGTTTCTACTATAGAAAAGCAGCAGTGGTGTTCACAAATCACTACAGAGCATTTCTG	2082	
Qy	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleLysValGln	1382	
Db	2083	TC TGCAAAACATCAAAACACAAGTCTATTTACAGATCAGAAGCAGTGTATCATTTATTCAA	2142	
Qy	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402	
Db	2143	GCTAGAAGTAAAGGATTTATACAGAACCGAAG-----TTTCAGGAAATTTAAA	2190	
Qy	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421	
Db	2191	AATAGCACCATAAAAATTCAGGCTATGCTGGAGAGATATAGACCAAGAAATATTATGT	2250	
Qy	1422	Lys-----AlaGlyTyrArg-----	1426	
Db	2251	AAAGTGAAGAGCTGCCTCAAGATTAACGCTGCTGTATAGATTGTTGGAGAGCACACAAGAA	2310	

QY	1426	-----	-----	-----	1426
Db	2311	TATCTAGCTATATATAAGCTGTATAAAATATTTCAGAGTTGCTTCTATATCAAACTAGAG	2370		
QY	1427	-----	-----	-----	1427
Db	2371	AGAACCGGTTTGAATGTGAGAGCATCAGCAATATATCATTCAGAGAAATCGAGGCT	2430		
QY	1442	ThrValGlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeu	1461		
Db	2431	ATACTTCTGCCAAAGATAGCTCATCAACACTTCTTAATGATAAAAGA	2478		
QY	1462	MetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMet	1481		
Db	2478	-----	-----	-----	2478
QY	1482	LysArgArgAlaAlaValValIleGlnArgPheArgAlaArgCysAlaMetLeuLys	1501		
Db	2479	---CATCGAGCTGCTTGTTCATCCCAAGCACATATAGAGGA	2523		
QY	1502	AlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArg	1521		
Db	2524	GGAAGGCAGCTCTTCTTCGGCAGAATCTGCTCTTGTATCATACAAAATATATACCA	2583		
QY	1522	AlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaVal	1541		
Db	2584	GCCAGGAGGCTGGAAGCATGAAGATAAATATATTTGAATTTAAANAATCTACAGTT	2643		
QY	1542	HisLeuGlnGlnLysPheArgGly-----	1558		
Db	2644	ATCCTACAGCACTGTGTGCTGTGTAGTACGAAAAAGATTTTATAGAACAGAGAGCC	2703		
QY	1559	CysPheGlnLeuLeuArg-----	1564		
Db	2704	AAATTCGACTCTTCACCTTCACTGCAGCTGCATATATATCACTGAAATGCTGTGAAT	2763		
QY	1564	-----	-----	-----	1564
Db	2764	CAAAAGACCTATAAACTTTACTTGGCTGTGAAGAATGCTAAACAGCAGGTTAATTCAGTC	2823		
QY	1565	---CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe	1583		
Db	2824	ATCTGTATTCAGAGATGTTTTTCAGACAAGA-----	2871		
QY	1584	---GlnAlaLeuMetThrProGluMetMetAsp-----	1596		
Db	2872	ATTCAGAAATATCATAGCATCAAAAAGATTGAGCATCAAGGTCAAGAAATGCTCGACCCAG	2931		
QY	1597	GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArg	1616		
Db	2932	CGAAATAGGCTGCATCAGTAATACAGAAAGCAGTCGCCATTTCTCTCCGTGATAAAG	2991		
QY	1617	GlnLysHis-----	1626		
Db	2992	CAGGAAAAATTCAGTAGTGAATCAITAAATTCAGGCATTATGAGAGCGCTATTCTTGG	3051		
QY	1627	-----	-----	-----	1627
Db	3052	AGGAAGAAAAATGATTGTACAAAAATTAAGCTATATACGATAAGTCTTCAAGTGTGTAAT	3111		
QY	1641	SerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSer	1660		
Db	3112	AGG-----	3122		
QY	1661	AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu	1680		
Db	3163	CAATTACCTTTTGACATATAAGACCTTCTTGCCATTCTTGAGGCGCTTAAACACCTAGAG	3222		
QY	1681	MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg	1700		
Db	3223	GTGTTACTAGA-----	3276		
QY	1701	SerGluValAspLysGlnLeuIleGluArgCysSerArg-----	1713		

Db	3277	ATTTCCTAAAAATATTTGTTTGTGATCCGAAGTGTGAATCGCAGTATTCTCTTGATGGAAAGTC	3336
QY	1714	-----llelleuAsnLeuAlaArgTyrAsnSerThrThrValAsn	1727
Db	3337	ATCAGATATGCTGTGCAAGTCTTGCTTAATGATATCTAAGTATGAGAAAACACTACTTCAGCA	3396
QY	1728	ThrPheGlnGluGlyLeuValThrIleAlaGlnMetLeuLeu	1742
Db	3397	GTTTATGATGTAGAAAATGTATAGATATACTATTGGAGCTTTTGCAGATATACCCAGAA	3456
QY	1743	-----ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeu	1758
Db	3457	AACCCCTGGTAATPAAAGTTGCAGCAAAAGGGGAAGCATTTTACAAAAAACTGTGTTG	3516
QY	1759	IleTrpValPheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsn	1778
Db	3517	TTGGCTATTTTACTG-----AAGACAACAATAGAGCCCTCTGATGTACGAAGTAGG	3567
QY	1779	ProGluAlaIleTyrMetValArgGluThrLysLeuValAlaArgLysGluLysMet	1798
Db	3568	TCCAAGTTGTTGACCGGTATTACAGCTCTCTCAAAACTTACAGCTCATTAACATAAAATG	3627
QY	1799	LysGlnAsnAlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys---	1817
Db	3628	-----AATACTGAAAGAATA-----CTTTACAAGCAAAAGAAGAT	3663
QY	1818	-----lleAsnPheThrProCysSer-----LeuProSer	1827
Db	3664	TCTTCTATAGCATTCCTTTTATCCAGAAACACCTGTAAAGACCAGAAATAGTTTCAAGA	3723
QY	1828	LeuGluProAspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyr	1847
Db	3724	CTTAAGCCAGATTGGGTTTGAGAAGAGATACATCGAAGAATCACAATCCCTGCGAA	3783
QY	1848	AlaPheAspThrIleLeuCysLysLeuGlnIle	1858
Db	3784	GCTATTCAATGGTATGGATAGCTTGGCAVTT	3816



Query Match:	7.44%	Indels:	319
DB:	13	Gaps:	40
US-09-914-698-1 (1-1861) x US-10-172-118-1746 (1-3887)			
Qy 868	TrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaIleThrValLeuGln	887	
Db 359	TGGATGATAGGAAAGAGTCGCA-----GAGATGCACAGGGCTGCTACTTTCATCCAG	412	
Qy 888	LysTrpTrpArgArgHisTrpLeuHis-----Val	897	
Db 413	TCTACTTTCAGAATGCACAGATTACATATGAGATATCAGGCTTTGAAACAGGCCCTCCGTT	472	
Qy 898	ValIleGlnArg-----ArgIleArgHisLysGlnLeuMet	909	
Db 473	GTGATCCAACAGCAATACCAAGCAAAATAGAGCTGCAAAATGCAGAGCGCAGCATATATCTC	532	
Qy 910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929	
Db 533	AGACAAAGACACTCTGCTGTGATCCTTCAGCTGCATTCAAGGTATGAAAACCTAGAGA	592	
Qy 930	TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThr	949	
Db 593	CAATTGAAGAGTATGCATTCC-----TCTGCAACCCCTATTTCAGAGTAGGTTT	640	
Qy 950	ArgArgTyrIleuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969	
Db 641	AGATCATCTCTGGTGAGGAGAAGATTCAATTCCTCCATAAAAGCTACTATTATTTGTTTCAG	700	
Qy 970	ArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg	989	
Db 701	AGGAAATATCGAGCCACCAATTTGCGCAACATAAATGTACCAATTTCTGCACTTAAGA	760	
Qy 990	GluAlaAlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeu---	1008	
Db 761	AAGCAGCCATTACATACAGCTCATCTTACAGAACTGTATGATAAGAGAAAGTTACAA	820	
Qy 1009	-----LeuAlaAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln---	1023	
Db 821	GAAATGCAAAAGGGCTGCAGT-TCTCATTTCAAGGCTACTTTCAAGATGCACAAAAA	879	
Qy 1024	-----GlnAlaAlaSerTyrIleGlnMetGlnTrpArgThr	1036	
Db 880	TATATTACATTTCAAGCTTGGAACATGCTTCAATTCATTTCAGCAACATTATCGAACA	939	
Qy 1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuIleMetPhe	1056	
Db 940	TATAGAGCTGCAAAATTCGAAAGAGAAATATATATCATGACAAATGGCATTTCTGCTTGGTT	999	
Qy 1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGluGlnArgLysGlu	1073	
Db 1000	ATTCAGCTGCATATAAAGGAATGAAAGCAAGACAACTTTTAAAGGGAATAAACACAAAGCT	1059	
Qy 1074	-----PheGlnGlnLeu	1077	
Db 1060	TCTATTGTAATACAAAGCCACTCAGAAATGTATAGGCAGTATTGTTTCTACAAAAGCTT	1119	
Qy 1078	LysArgAlaAlaIleAlaIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097	
Db 1120	CAGTGGGCTCAAAAATCATACAGAAAAATATAGACAAATATAAAGAAACAGAAA---	1176	
Qy 1098	AsnAlaAspTyrIleuAlaLeuArgSerSerValLeuLys-----ValGlnAla	1113	
Db 1177	-----GTATTTCAACAATGAATTAAGAAAGAGACTTGTGTTTCAGGCA	1221	
Qy 1114	---TyrArgLysAlaThrIleGlnMetArgIleAsp-----	1124	
Db 1222	GGTTTTTCAGGCATGAACATAAATAAACAGATTTCAGGAACAGCCAGGCTGCCATTATT	1281	
Qy 1125	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1134	
Db 1282	ATTCAAGACATTGTAAAGCCTTTAAATAATAGGAAGCAATTATCTCCACATTAGAGCAACA	1341	

Qy	1135	ValleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	1151
Db	1342	GTAGTTTCTATTCAAGAAGATACAGAAAACAACTACTCGTGGTACCCACGAGTTATT	1401
Qy	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
Db	1402	TGTATACAGTCTTATTACAGAGCTTTAAAGTACGAAGGATATTCAAATATGACCGG	1461
Qy	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
Db	1462	GCTGCCACACTAATTCAGTCTATTCTCATGCAATGCACAGGCCCAAGTTGAT-----	1512
Qy	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
Db	1513	---TATGAAACAAAGAAAACCTGCAATTTGTTGTTATACAGAAATTATTATAGTGTGTATGTT	1569
Qy	1201	GlnMetArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGln	1220
Db	1570	AGAGTAAAAACAGAAAGAAAACCTTTTAGCAGTTCAGAAATCTGPACCAACTATTTCAG	1629
Qy	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLys	1240
Db	1630	GCTGCTTTTAGAGGCATGAA-----GTTAGACAAAATTTGAAAATGATCAGAGGAA	1683
Qy	1241	LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
Db	1684	AAGATGCAGCCATTGTT-----AACCAATCTGCACCTCTGCTGTTACAGAAGTAA	1734
Qy	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgGpHeHisAlaGln	1280
Db	1735	ACTCAGTATGAAGCTGTTCAAAGTGAAGGTGTTATGATTCACAGAGTGGTATAAGCTTCT	1794
Qy	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeu	1300
Db	1795	GGCCTTGCTGTTCCAGAGACGAGTATCATCTCAAAGTAGGGCTGCAGTAAACATT	1854
Qy	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
Db	1855	CAAAAAGCTTTTGTAGAATGGTCACAGAAAACCTGMAAACACAGAAAATGCTGCTGCCCTA	1914
Qy	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
Db	1915	CGGATTCAGTCTCTCCCTTCAGATGGCTGTGTATCGGAGAGATTGTTCCAGCAGAAAAGA	1974
Qy	1323	AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLys	1342
Db	1975	GCTGCTATCACTTTACAGCATTTATTTTAGACG-----TGGCAAAACCGAANA	2022
Qy	1343	SerTyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
Db	2023	CAGTTTCTACTATAGAAAAGCAGCAGTGGTGTTTTACAAAATCACTACAGCATTTCTG	2082
Qy	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
Db	2083	TCGTGCAAAACATCAAGACAAAGTCTATTACAGATCAGAGCAGTGTATCAATTATTCAA	2142
Qy	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
Db	2143	GCTAGAGTAAGATTTATACAGAAACGGAAG-----TTTCAGGAAATTTAA	2190
Qy	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
Db	2191	AATAGCACCATAAAAATTCAGGCTATGTGGAGAGATATAGAGCCCAAGAAATATTATGCT	2250
Qy	1422	Lys-----AlaGlyTyrArg-----	1426
Db	2251	AAAGTGAAGCTGCCTGCAAGATTCAAGCCTGTGTATAGATGTTGGAGAGCACAAAGAA	2310
Qy	1426	-----	1426
Db	2311	TATCTAGCTATATTAAGAGCTGTAAATATTATTCAGGTGCTTCTATATCCAACTAGAG	2370
Qy	1427	-----ArgIleArgLeuSerSerIleGlnArgLysTrpArgAla	1441

2371	Db	AGAACACGGTTTTTGAATGTGACGAGCATCAGCAATTATTCATTCAGAGAAATCGAGAGCT	2430
1442	Qy	ThrValGlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeu	1461
2431	Db	ATACTTCTCGAAGATAGCTCATGAACACTTCTTAATGATAAAAGA	2478
1462	Qy	MetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMet	1481
2478	Db	-----	2478
1482	Qy	LysArgGlnAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLys	1501
2479	Db	---CATCGAGCTGCTGTGTTGATCCAAAGCACATTATAGAGA-----TATAAA	2523
1502	Qy	AlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArg	1521
2524	Db	GGAAGGCAGGCTCTTCTTCGCGACAAATCTGCTGTTTGATCATACAAAATATATACGA	2583
1522	Qy	AlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaVal	1541
2584	Db	GCCAGGAGGCTGGAAGCATGAAGAGATAATATATTAATGAATTTAAAAAATCTACAGTT	2643
1542	Qy	HisLeuGlnGlnLysPheArgGly-----LysArgLeuMetIleGluGlnArgAsn	1558
2644	Db	ATCCTACAAGCACTGGTGCCTGGTGGCTAGTACGAAAAAGATTTTATAGAACAGAGGCC	2703
1559	Qy	CysPheGlnLeuLeuArg-----	1564
2704	Db	AAATTCGACTCTCTCACTTCACTGACGCTGCATATTATCACTGAAATGCTGTAGAATT	2763
1564	Qy	-----	1564
2764	Db	CAAGACGCTATAAACCTTTACCTGGCTGTGAAGATGCTCAACAGCAGGTAAATTCAGTC	2823
1565	Qy	---CysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe	1583
2824	Db	ATCTGTATTACAGATGCGTTTCGAGCAAGA-----TTACAAGAAAAAGAGATTT	2871
1584	Qy	---GlnAlaLeuMetThrProGluMetMetAsp-----LeuIleArg	1596
2872	Db	ATTCAGAAAAATCATAGCATCAAAAAGATTGAGCATGAAGGTCAAGAAATGCTTGAGCCAG	2931
1597	Qy	GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArgArgArg	1616
2932	Db	CGAAATAGGCTGCATCAGTAATACAGAAAGCAGTGCCTATTTCTCTCCCGTAAAAAG	2991
1617	Qy	GlnLysHis-----GlnGlyLeuLeuAspIleArg-----	1626
2992	Db	CAGAAAAATTCACTAGTGGAAATCAATAAAATTCAGCATATGAGGAGGAGCTATTCCTGG	3051
1627	Qy	-----LysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsn	1640
3052	Db	AGGAAGAAAAATGATTGTACAAAAATTAACGATATACGCTAGTCTTCAAGTTGTTAAT	3111
1641	Qy	SerValArgCysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheIleAlaSer	1660
3112	Db	AGG-----GAGATTTCGAGAGAAAAACAACTCTACAAAAGAACTGCACCTTCGCACTT	3162
1661	Qy	AspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeu	1680
3163	Db	CAITACCTTTTGACATATAGACACTTTCGCGCAATCTTGAGGCTTTAAACACCTTAGAG	3222
1681	Qy	MetTrpCysSerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArg	1700
3223	Db	GTAGTTACTAGA-----TTGCTCCCACTTTGTTGTGAGAACATGGCCAGAGTGGAGCA	3276
1701	Qy	SerGluValAspLysGlnLeuIleGluArgCysSerArg-----	1713
3277	Db	ATTCTCAAAATATTGTTTGTGATCCGAAGTTGTAATCGAGTATTCGAGTATTCCTGTATGGAAGTC	3336
1714	Qy	-----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsn	1727

## RESULT 5

```

US-10-115-635-64
; Sequence 64, Application US/10115635
; Publication No. US20040137434A1
GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Dрманac, Radcoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: polypeptides
; FILE REFERENCE: 797CON
; CURRENT APPLICATION NUMBER: US/10/115,635
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 09/714,936
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 362
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 64
; LENGTH: 4027
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1021)..(3825)
US-10-115-635-64

Alignment Scores:
Pred. No.: 1.53e-58
Score: 700.50
Percent Similarity: 40.70%
Length: 4027
Matches: 275
Conservative: 226

```

Best Local Similarity:	22.34%	Mismatches:	412
Query Match:	7.36%	Indels:	319
DB:	17	Gaps:	40
US-09-914-698-1 (1-1861) x US-10-115-635-64 (1-4027)			
QY	868	TyrGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAlaThrValLeuGln	887
DB	359	TGGATCATAGGAAAGGATGCGA-----GAGATGCCAGGGCTGCTACTTTTCATCCAG	412
QY	888	LysTrpTyrArgGHisTrpLeuHis-----Val	897
DB	413	TCTACTTTTTCAGATGCACAGATTACATATACAGATATCAGGCTTTGAAACAGGCGCTCCGTT	472
QY	898	ValIleGlnArg-----ArgIleArgHisLysGluLeuMet	909
DB	473	GTGATCCCAACAGCAATPACCAAGCAATAGAGCTGCAAAACTGCAGAGCGAGCATTTATCTC	532
QY	910	ArgArgHisArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLys	929
DB	533	AGCAAAAGACATCTGCTGTATCTTCAGGCTGCAATTCAGGGTATGAAACTAGAGAA	592
QY	930	TyrValLysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThr	949
DB	593	CATTTCAGAGATGATGATCC-----TCTGCAACCCCTTATTCAGAGTAGGTTT	640
QY	950	ArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGln	969
DB	641	AGATCATTTACTGGTGAGGAGAGATTCATTTCCCTCAAAAAGCTACTATTTTGTTCAG	700
QY	970	ArgTyrTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArg	989
DB	701	AGCAATATTCGAGCCACCATTTGTGCAACATATAAATTTGATACCAATCTTGCACTTAAGA	760
QY	990	GluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeu---	1008
DB	761	AAGCGACCATTTACATACAGTCATCTTACAGAGACTGTGCTAAAGAGAGTTTACAA	820
QY	1009	-----LeuAlaAlaGluThrAlaArgLeuGlnArgSerGlnLysGln---	1023
DB	821	GAATGCCAAAGGCTGCACT-TCTCATTCAGGCTACTTTCCAGATGCACAGAAAAA	879
QY	1024	-----GlnAlaAlaAspTyrIleGlnMetGlnTrpArgThr	1036
DB	880	TATATTACATTTTCAGACTTGGAAACATGCTTCAATTCATTTTCAGCAACATTTACGACA	939
QY	1037	TyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAspLeuMetPhe	1056
DB	940	TATAGAGCTGCAAAATTCAGAGAGAAATTTATATCAGACAAATGCGCATTTCTGCTGGTT	999
QY	1057	ValGln-----ArgArgMetArgSerLysTrpSerMetLeuGlnArgLysGlu	1073
DB	1000	ATTTCAGGCTGCATATAAAGGAATGAAAGCAAGCAACTTTTAAGGGAACAAACAAAGCT	1059
QY	1074	-----PheGlnGlnLeu	1077
DB	1060	TCTATCGTAAATCAACAGCACCTACAGAAATGTATAGCAGTATTGTTCTACCAAAAGCTT	1119
QY	1078	LysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSerMetArgLysCys	1097
DB	1120	CAGTGGCTTACAAAATCATACAGAAAAATATAGAGCAAAATAAAGAAACAGAAAA---	1176
QY	1098	AsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLys-----ValGlnAla	1113
DB	1177	-----GTATTTTCAACAATGAACTTAAGAAAGAGAGACTTTGTGTTTCAGGCA	1221
QY	1114	---TyrArgLysAlaThrIleGlnMetArgIleasp-----	1124
DB	1222	GGTTTTTCAGACATGAACATATAAAGAAACAGATTCAGGAACAGCAGCGCTGCCATTATT	1281
QY	1125	-----ArgAsnHisTyrTyrSerLeuArgLysAsn	1134
DB	1282	ATTTCAGAGCATTTGTAAGGCTTTTAAATATAAGGAGCATTTATCTCCACCTTAGAGCAACA	1341

QY	1135	ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGln-----	1151
DB	1342	GTAGTTTCTTATTTCAAAGAGATACAGAAAACTACGTGCGTGCCTACCCACAGAGTTATT	1401
QY	1152	-----ArgGluAsnTyrLeuArgLeuArgAsn	1160
DB	1402	TGTATACAGCTCTTATTTACAGAGGCTTTAAAGTACGAAAGGATATTCAAATATGCAACCG	1461
QY	1161	AlaSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsn	1180
DB	1462	GCTGCCACACTAATTCAGTCAATTCATGCAATGCACAGCGCCCAAGCTTGTAT	1512
QY	1181	AlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThrLeu	1200
DB	1513	---TATGAAACAAAGAAAACTGCAATTTGGTTATACAGAAATTTATAGTTGTATGTT	1569
QY	1201	GlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysArgIleGln	1220
DB	1570	AGAGTAAAAAAGAAAAAATTTTTCAGAGTTTCAGAAATCTGTACGAACATTTCTAG	1629
QY	1221	IleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLys	1240
DB	1630	GCTGCTTTTTCAGAGGATGAA-----GTTAGACAAAAATTCAAAAATCTATCAGAGGAA	1683
QY	1241	LysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArg	1260
DB	1684	AGATGGCGACCATTTGTT-----AACCAATCTGCTCTCTGCTGTACAGAAACAA	1734
QY	1261	GlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGln	1280
DB	1735	ACTCAGTATGAAGCTTTCAAAGTGAAGGTGTATGATTCAGAGTGGTATATAAGCTTCT	1794
QY	1281	LysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeu	1300
DB	1795	GGCTCTGCTGTTTCAGAGGAGCAGAGTATCATTTCTCAAAGTAGGCTGCTGCAACAAT	1854
QY	1301	GlnMetHisTrpArgAsnHisLeuLeuArgLysArgGlu-----	1313
DB	1855	CAAAAGCTTTTGTAGAAATGTCACAGAAAACTGGAAACACAGAAATGTGTCGCCCTA	1914
QY	1314	-----ArgAsnSerPheLeuGlnLeuArgGln	1322
DB	1915	CGGATTCAGTTCTCTCTTCAGATGGCTGTATCGGAGAGATTTGTTTCAGCAGAAAAA	1974
QY	1323	AlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLysLys	1342
DB	1975	GCTGCTATCATCTTTCAGCATTTATTTTAGGAGG-----TGGCAAAACAGAAAA	2022
QY	1343	SerTyrAlaGlnLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLys	1362
DB	2023	CAGTTTCTTACTATAGAAAAAGCAGCGTGGTTTTCAAAAATCACTACAGAGCTTTCTG	2082
QY	1363	AlaMetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGln	1382
DB	2083	TCTGCAAAACATCAAGACAGCTATTTTACAGATCAGAGCAGTGTATTCATTATTCAA	2142
QY	1383	ArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArg	1402
DB	2143	GCTAGAGTAAAGGATTTTATACAGAAACCGAAG-----TTTCAGGAAATTTAAA	2190
QY	1403	GlnAlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArg---LeuCys	1421
DB	2191	ATAGCACCATAAAAATTCAGGCTATGTGGAGAGATATAGAGCCCAAGAAATTTTATGT	2250
QY	1422	Lys-----AlaGlyTyrArg	1426
DB	2251	AAAGTGAAGCTGCCTGCAAGATTCAAGCTCGTATAGATGTTGGAGAGCACACAAGAA	2310
QY	1426	-----	1426
DB	2311	TATCTAGCTGTATTAAGCTGTATAAATTTATTTCAAGGTTGCTTCTATACCAAACTAGAG	2370



Db 131 AAACAATATTTCATTTAAAGACGAGCTGCTACTATTATTCAGCAATGT----- 178  
 Qy 745 ThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThr 764  
 Db 179 -----TACCGTTCCAAATAATAGCT----- 199  
 Qy 765 ArgGluLeuArgGluLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPhe 784  
 Db 200 -----GCACAAAGAGAGAGATATATG----- 223  
 Qy 785 AspTyrAlaPheAsnAsnLeuAlaValAspLeuArgAspGly---ValArgLeuThrArg 803  
 Db 224 -----CAGATCGCGAATCTTGATCAATCAATCGCAAGCA 256  
 Qy 804 ValValGluValIleLeuLeuArgAspLeuThrArgGlnLeuArgValProIle 823  
 Db 257 TTTGTTAGAGGATACCTTCGGAAGCAGATG-----AGGTTACAAAGAAAGCTGTT 310  
 Qy 824 SerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGlyAlaLeuGlyGluAlaAsn 843  
 Db 311 ATTTCACTACAGTCTATTTCAGATGAGAAAGGCT----- 346  
 Qy 844 PheGlnLeuGlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThr 863  
 Db 347 -----CGGCAATATTAT 358  
 Qy 864 LeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHisAlaAla 883  
 Db 359 CTG-----AAATGTATAAGCAATT 379  
 Qy 884 ThrValLeuGlnLysTrpTrpArgHisTrpLeuHisValValIleGlnArgGile 903  
 Db 380 ATTGTCATTCAAGTACTAT-----CATGCATACAAAGCACAGGTCAATCAGAGG--- 430  
 Qy 904 ArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheArg 923  
 Db 431 -----AAGAACTTCTTGCAAGTCAAAAAGCAGCTACTTGTGTCAGCAGCTTACAGA 484  
 Qy 924 GlyHisGlnMetArgLysTyrValLysLeuPheLysThrGluArgThrGlnAlaAla 943  
 Db 485 GGTATAAGTACGCGCAGCTAATCAAA-----CAACAATCTATAGCTGCTCTT 532  
 Qy 944 IleLeuGlnLysPheThrArgArgTyrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHis 963  
 Db 533 AAAATTCAGTCTGCTTTAGAGGCTATAATAAAGCGTAAATATATCAATCTGCTTCAA 592  
 Qy 964 SerIleIleThrIleGlnArgTrpTrpArgAlaGlnGlnLeuGlyArgGlnHisArgGln 983  
 Db 593 TCTATAATAAGATTTCAGAGATGTTACAGGCGGTACAGAGACTCTTCATCATACAAAGCA 652  
 Qy 984 ArgPheValGluLeuArgGluAlaAlaIlePheLeuGlnArgIleTrpArgArgLeu 1003  
 Db 653 CATTTTTTGAAGCAAAAGCAGCTGTGATTTCCCTCCAGTCTGCTTATCGTGCTGGAAG 712  
 Qy 1004 PheAlaLysLysLeuAlaAlaAlaGluThrAla-----Arg 1016  
 Db 713 GTTCGGAACACATTAGAGGGAACATCAAGCTGCTTGAAGATTTCAGTCTGCTTTTGA 772  
 Qy 1017 LeuGlnArgSerGlnLysGln-----GlnAlaAlaSerTyrIleGlnMet 1032  
 Db 773 ATGGCCAGGCCAGAAACAGTTTAGATTGTTTAAACAGCAGCAGCATTAGTCTCCAGCAA 832  
 Qy 1033 GlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnArgAsp 1052  
 Db 833 AATTTCAGAGCATGGACTCCAGGAAGGAAGCAATGTATGGAGTATATTGAATCCGTCAT 892  
 Qy 1053 LeuIleMetPheValGln-----ArgArg----- 1060  
 Db 893 GCGGTACTGCTTCAATCTATGTGGAAGGGAACACACTGAGAGACAGCTTCAAGG 952  
 Qy 1061 -----MetArgSerLysTrpSerMetLeuGlnArgLysGlu 1073  
 Db 953 CAACATAAATGTGCTATCATCATACAGTCTATATAGATGATGATGTCGCAACAAAGAGAG 1012

Qy 1074 PheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArgAlaLysLeuSer 1093  
 Db 1013 TGGAAAAATCAATAAAAAAGCTGCTCTCTGATTCAAGAGTATTATAGGCTTACAGTATT 1072  
 Qy 1094 MetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeuLysValGln--- 1112  
 Db 1073 GGAAGAGACAGACATCAATTATTTGAAACAAAGACAGCTGTAGTAATTACAGTCA 1132  
 Qy 1113 AlaTyrArgLysAlaThrIleGlnMetArgIle----- 1123  
 Db 1133 GCTTATCGTGGTATGAAAGTGAAGAAATATAAAGATTGCAACAAAGCAGCAGTCACT 1192  
 Qy 1124 -----AspArgAsnHisTyrTyrSerLeuArgLysAsn 1134  
 Db 1193 ATACAGTCTAAATACAGAGCTTACAAACCAAAAGAAATATGCACCTATAGAGTTCA 1252  
 Qy 1135 ValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArgGluGlnArgGluAsn 1154  
 Db 1253 GCTATTATAATTACAGATGTTATCGAGTATTAATAATTACAAACCATCAGCATAAGGAG 1312  
 Qy 1155 TyrLeuArgLeuArgAsnAlaSerIleLeuValGlnLysArgTyr-----ArgMet 1171  
 Db 1313 TATCTTAATTTGAGAGACAGCAATTAATAATCCAAATCTGTTTATAGAGGTATTAGATT 1372  
 Qy 1172 ArgGln-----GlnMet 1175  
 Db 1373 AGAAGACATATTCAACATGACAGGCGGAGCCACTTTTATTAAAGCCATGTTAAAAATG 1432  
 Qy 1176 IleGlnAspArgAsnAlaTyrLeuArgThrArgLysCysIleIleAsnValGlnArgArg 1195  
 Db 1433 CATCAGTCAAGATAAGTTTACCATACATACAGAAAGCAGCTATTGTTTATCAGTAAAG 1492  
 Qy 1196 TrpArgAlaThrLeuGlnMetArgGluArgLysAsnTyrLysLeuHisLeuGlnThrThr 1215  
 Db 1493 TGTAGAGCATATTATCAAGGTAAATGCGCGGTGAAAGTACCTGACAATTTTGAAGCT 1552  
 Qy 1216 ThrLysArgIleGlnIleLysPhe-----ArgAlaLysArgGluMetLysLys--- 1231  
 Db 1553 GTTAAAGTCTTTCAGGCAAGTTTTAGAGGAGTAAAGATTAGACGGACTTTAGAAAGATG 1612  
 Qy 1232 -----GlnArgAlaGluPhe 1236  
 Db 1613 CAGACTGCGACCAACTCATTCAGTCAAACTACAGAGATACAGACAGCAACATACITTT 1672  
 Qy 1237 LeuGlnLeuLysLysValThrLeuValValGlnLysArgArgArgAlaLeuLeuGlnMet 1256  
 Db 1673 AATAAGTTAAAGAAATAACAAAAACAGTACAGCAAGATACCTGGGCAATGAAAGAGA 1732  
 Qy 1257 ArgLysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArg 1276  
 Db 1733 AACATACAAATTCGAAAGGTATTAACAAACTGAGGCAATTCGTAAATATACATTCAGGCTATT 1792  
 Qy 1277 PheHisAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAla 1296  
 Db 1793 TTTAGGGAAGAAAGCTAGAGACATTTAAAAATGATGCAT-----ATAGCC 1840  
 Qy 1297 ValSerCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLysArgGluArgAsnSer 1316  
 Db 1841 GCAACTCTCATTCAGAGGAGATTAGAACTCTTAATGATGAGAAGAAGA----- 1888  
 Qy 1317 PheLeuGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsn 1336  
 Db 1889 TTCTCTCTCTCAGAAAACTGCTATTGTTGATTCGAGAAAAATATCGGGCAATCTTT--- 1945  
 Qy 1337 MetIleLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThr 1356  
 Db 1946 TGTAACAAGCATCACTTACAGTTCCTTCAGGTACAAAATGCAAGTTTAAAAATCCAGTCA 2005  
 Qy 1357 ArgTyrArg-----AlaLysLysAlaMetGlnLysGln----- 1367  
 Db 2006 TCATACAGAAAGATGATGATAAGGAAGGATGCGAGAGATGCACAGGCGTCTACTTTC 2065

QY 1368 -----ValValLeuTyrGlnLysGlnArgGluAla 1377  
 Db ATCCAGTCTACTTTTCAGAAATGACAGATATCATATGAGATATCAGGCTTTGAAACAGAGCC 2125  
 QY 1378 IleIleLysValGlnArgArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluVal 1397  
 Db TCCGTTGTGATCCACAGCAATACCAAGCAATAGAGCTGCAAACTGCGAGGCGACAT 2185  
 QY 1398 TyrGlnLysGlnArgGlnAlaValIleArgLeuLysTrpTrpArgSerIleArgAsp 1417  
 Db TAICTCAGACAAAGACACTCTGCTGTGATCTTCAGGCTGCATTCAGGGGTATGAAACT 2245  
 QY 1418 MetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerIleGlnArg 1437  
 Db AGAAGACATTTGAAGACT-----ATGCATTCTCTGCAACCCCTATTTCAGAGT 2293  
 QY 1438 LysTrpArgAlaThrValGlnAlaArgArgGlnArgGluLeuPheLeuSerThrIleArg 1457  
 Db AGGCTTAGATCATTTACTTGGTGCAGGAGAGA-----TTCAATTCCTCCCAAAAA 2341  
 QY 1458 LysValArgLeuMetGlnAlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArg 1477  
 Db GCTACTATTTTGTTCAGAGAAATATCGAGCCACCAATTTGTGCCAAACATAAATTTGATC 2401  
 QY 1478 GluPheGluMetLysArgArgAlaAlaValIleGlnArgArgPheArgAlaArgCys 1497  
 Db CAATTCCTGCACCTTAAGAAAGGCGACCATTAACAATACAGTCATCTTACAGA----- 2452  
 QY 1498 AlaMetLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGln 1517  
 Db ---AGACTGATGTTAAAGAAAGAGTTTACAGAAATGCAAAAGGCTGCAGTTCATTCAG 2509  
 QY 1518 ArgLysPheArgAlaAsnArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArg 1537  
 Db GCTACTTTCAGTGCACAGACA-----TATATTACATTTTCAG 2548  
 QY 1538 Thr-----IleAlaValHisLeuGlnGlnLysPheArgGlyLysArgLeuMetIle 1554  
 Db ACTTGGAAACATGCTTCAATTCATTAATTCAGCAACATTTATGCAATATAGAGCTGCAAAA 2608  
 QY 1555 GluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheGlnAlaArgAla 1574  
 Db TTGCAAGAGAGAA----- 2620  
 QY 1575 ArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGluMetMetAspLeu 1594  
 Db -----AATTAT 2626  
 QY 1595 IleArgGlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrLeuIleArg 1614  
 Db ATCAGACAATGGCATCTGCTGTGTTATTTCAGGCTGCATATAAGGAATGAAAGCAAGA 2686  
 QY 1615 Arg-----ArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeu 1632  
 Db CAACTTTTAAGGGAACCAAGCTTCTATCTGTAATACAAAGCACTCAGAAATGTAT 2746  
 QY 1633 ArgGlnGluAlaLysAlaValAsnSerValArgCys-----LysValGlnGluAla 1649  
 Db AGGAG-----TATTTGTTTCTACCAAAAGCTTCAGTGGCT 2782  
 QY 1650 ValArgPheLeuArgGlyArgPhe 1657  
 Db ACAAAATCATACAAAGAAAAATAT 2806

RESULT 7

US-10-342-887-1748  
 ; Sequence 1748, Application US/10342887  
 ; Publication No. US20040058340A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter S.  
 ; APPLICANT: Mao, Mao

; APPLICANT: Roberts, Christopher J.  
 ; APPLICANT: Van 't Veer, Laura Johanna  
 ; APPLICANT: Van de Vijver, Marc J.  
 ; APPLICANT: Bernards, Rene  
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
 ; FILE REFERENCES: 9301-188-999  
 ; CURRENT APPLICATION NUMBER: US/10/342,887  
 ; CURRENT FILING DATE: 2003-01-15  
 ; PRIOR APPLICATION NUMBER: 60/298,918  
 ; PRIOR FILING DATE: 2001-06-18  
 ; PRIOR APPLICATION NUMBER: 60/380,710  
 ; PRIOR FILING DATE: 2002-05-14  
 ; PRIOR APPLICATION NUMBER: 10/172,118  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 2699  
 ; SEQ ID NO 1748  
 ; LENGTH: 2414  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-342-887-1748

Alignment Scores:  
 Pred. No.: 1,32e-32 Length: 2414  
 Score: 437.00 Matches: 210  
 Percent Similarity: 36.74% Conservative: 153  
 Best Local Similarity: 21.26% Mismatches: 331  
 Query Match: 4.59% Indels: 294  
 DB: 13 Gaps: 33

US-09-914-698-1 (1-1861) x US-10-342-887-1748 (1-2414)

QY 913 ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys 932  
 Db 2 AAATCTGACGAATATTCAGGCTGCTTTAGAGGATGAAGTTAGACAAAATTGAAA 61  
 QY 933 LeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyr 952  
 Db AATGTATCAGAGGAAAAG---ATGGCAGCCATTGTTAACCAATCTGCACCTCTGCTGTATC 118  
 QY 953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr 971  
 Db AGAAGCAAAACTCAG---TATGAAGCTGTTCAAGTGAAGTGTATGATTCAGAGTGG 175  
 QY 972 TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla 991  
 Db TATAAGCTTCGSCCTTGTCTGTTTCACAGAGACAGATATCATTTCAAAGTAGGCT 235  
 QY 992 AlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLysLysLeuLeuAla 1011  
 Db GCAGTACAAATTCAAAAGCTTT---TGTAGAATGGTCAAGAACTG----- 283  
 QY 1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaAlaSerTyrIleGln 1031  
 Db 284 -----GAAACACAGAAATGTGCTGCCCTCCATCGGATTCAG 316  
 QY 1032 -----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg 1049  
 Db TTCTTCCTTCAGATGGCTGGTAT-----CGGAGAACATTTGTTTCAG 358  
 QY 1050 GlnArgAspLeuLeuMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlu 1069  
 Db CAGAAAAGAGCTGCTATCATCTTTACAGCATTTATTTAGGACG---TGG-----CAA 406  
 QY 1070 GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTyrArg 1089  
 Db 407 ACCAGAAAACAGTTTTTACTATATAGAAAAGCAGCAGTGGTTTTACAAAATCATCTACAGA 466  
 QY 1090 AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu 1109  
 Db GCATTTCTGTCTGCAAAACATCAAGACAAAGTCTATTATTCAGATCAGACAGCAGTGTATC 526  
 QY 1110 LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyr 1129

Db 527 ATTATTCAAGCTAGAGTAAGATTATACAGAAACGG----- 565  
Qy 1130 SerLeuArgLysAsnValIleCysLeuGlnArgLeuArgAlaIleMetLysMetArg 1149  
Db 565 ----- 565  
Qy 1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln----- 1166  
Db 566 -----AAGTTTCAGAGAAATTAATAATAGCACCATAATAATTCAGGCTATGG 613  
Qy 1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186  
Db 614 AGGAGATATAGACCAAGAAA-----TATTATGTAAGTG 649  
Qy 1187 LysCysIleIleAsnValGln-----ArgArgTyrArgAlaThrLeuGlnMetArg 1203  
Db 650 AAAGCTGCTGCAGATTCAAGCTGGATAGATGTGGAGAGCA----- 694  
Qy 1204 ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPhe 1223  
Db 695 -----CACAAAGATATCTAGCTGTATTAAGCTGTAAATATTATCAAGTTGCTTC 748  
Qy 1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLysValThr 1243  
Db 749 TATACCAAACTAGAG-----AGAACACGGTTTTTGAATGTGAGAGCATCAGCA 796  
Qy 1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263  
Db 797 ATTATCATTCAGAGAAATGAGAGCTATCTCTGCAAGATAGCTCATCAACACTTC 856  
Qy 1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAlaGlnLysSerMet 1283  
Db 857 TTAATGATAAAAGA----- 871  
Qy 1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValSerCysLeuGlnMetHis 1303  
Db 872 -----CATCGAGCTGCTTGTGTGATCCCAAGCACAT 901  
Qy 1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323  
Db 902 TATAGAGCA----- 910  
Qy 1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343  
Db 911 -----TATAAGAGGCGAGGCTCTCTCTCGGAG----- 940  
Qy 1344 TyrAlaGlnLeuLysGlnAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363  
Db 941 -----AAATCTGCTGTGATCATACAAAAATATATACGAGCCAGCGAGGCT 988  
Qy 1364 MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg 1383  
Db 989 GGAAGCATGAAGAGTAAATATATTGAATTTAAATAATCTACAGTTATCTCTACACAGCA 1048  
Qy 1384 ArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGln 1403  
Db 1049 CTGTGCTGCTGTGCTAGTACGAAAGATTTTAAAG-----CAGAGA 1093  
Qy 1404 AlaValIleArgLeuGlnLysTrpTrpArgSerIleArgAspMetArgLysLysAla 1423  
Db 1094 GCCAAAATTCGACTT-----CTTCACCTCACTGCAGCT 1126  
Qy 1424 GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal 1443  
Db 1127 GCATATTATCACTGAAT----- 1144  
Qy 1444 GlnAlaArgArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
Db 1144 ----- 1144  
Qy 1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgArgGluPheGluMetLysArg 1483  
Db 1144 ----- 1144

Qy 1484 ArgAlaIleValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
Db 1145 -----GCTGTGTAGAAATCAAGAGCCTATATAAATTTTACCTGGCTGTGAAGATGCTAAC 1198  
Qy 1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
Db 1199 AAGCAGGTT-----AATTCAGTCATCTGTATTCAGAGATGGTTTCGAGCAGCA 1246  
Qy 1524 ArgSerMetLysGlnAlaArgGlnPheValGlnLeuArgThrIleAlaValHisLeu 1543  
Db 1247 TTACAAGAAAAGAGATTTATTTCAGAAATATCATAGCATC----- 1285  
Qy 1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeu 1563  
Db 1286 -----AAAAGATTGAGCATGAAGGTCAAGATGTCTGAGCCAGCGCA 1327  
Qy 1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
Db 1328 AATAGGGCTGCATCAGTAATACAGAAAGCAGTGGCCATTTTCTCTCCGTAAAAAGCAG 1387  
Qy 1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaLysVal 1603  
Db 1388 GAAAAAATTCACCTAGT-----GGAATCATATAA----- 1414  
Qy 1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeu 1623  
Db 1415 ATTCAGGCATATGAGAGAGCTTATTTCTGGAGAGAAAGAAATGATGTACAAAATATAA 1474  
Qy 1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg 1643  
Db 1475 GCTATACGACTAAGTCTTCAAGTGTAAATAGGAGATTCGAGAGAAAACAACTCTTAC 1534  
Qy 1644 CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663  
Db 1535 AAAAGAACTGCACCTTGCACTTCACTTTTGCATATATAAGCACCTTTCTGCCATCTT 1594  
Qy 1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1683  
Db 1595 GAGGCTTTAAACACCTAGAGGTAGTTACTAGA----- 1627  
Qy 1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703  
Db 1628 -----TTGTCTCCACTTTGTGTGAGAACATGGCCGAGAGTGGAGCAATTTCTAAA 1678  
Qy 1704 AspLysGlnLeuIleGluArgCysSerArg----- 1713  
Db 1679 ATATTGTGTTGATCCGAAAGTTGTAATCGCAGTATTCCTTGTATGGAAGTCATCAGATAT 1738  
Qy 1714 -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
Db 1739 GCTGTGCAAGTCTTGTCTTAATGTATCTAAGATGAGAAAACACTTCTCAGCAGTTTATGAT 1798  
Qy 1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
Db 1799 GTAGAAAATTTATAGATATATCTATTGGAGCTTTTGAGATATACCGAGAAAAGCTGT 1858  
Qy 1743 ---ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrIleTrpVal 1761  
Db 1859 AATAAAGTTGCAGACAAAGCGGAGCATTTTACAAAAACTTGTGTTGTTGTTGCTATT 1918  
Qy 1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781  
Db 1919 TTACTG-----AAGACAACAATAAGACGCTCTGTATGACGAGTAGGTGCCAAGTT 1969  
Qy 1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801  
Db 1970 GTTGACCGTATTACAGTCTCTACAACTTACAGCTCATAAACATAAATG-----AAT 2023  
Qy 1802 AlaArgLysProProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817  
Db 2024 ACTGAAAGATA-----CTTTACAGCAAAAGAGAAATTTCTTCTATA 2065



```
QY 1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2066 AGCATCTCTTTATCCAGAAACACCTGTGAAGCACCAGATAGTTTCAAGACTTAAGCCA 2125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2126 GATTGGGTTTGAAGAGATAAATGGAAGAAATCAAAATCCCTGCAAGCTATTCAA 2185
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1851 ThrIleLeuCysLysLeuGlnIle 1858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2186 ATGGTGATGATACGCTTGCCATT 2209
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-10-172-118-1748
; Sequence 1748, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1748
; LENGTH: 2414
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_018136
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1748

Alignment Scores:
Pred. No.: 1.32e-32 Length: 2414
Score: 437.00 Matches: 210
Percent Similarity: 36.74% Conservative: 153
Best Local Similarity: 21.26% Mismatches: 331
Query Match: 4.59% Indels: 294
DB: 13 Gaps: 33

US-09-914-698-1 (1-1861) x US-10-172-118-1748 (1-2414)
QY 913 ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys 932
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 AAATCTGTAGAACTATTACGGCTGCTTTAGAGCATGAAGTGTAGACAAAATTGAAA 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 933 LeuPheLysThrGluArgThrGlnAlaAlaIleIleLeuGlnLysPheThrArgArgTyr 952
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 AATGTATCAGAGGAAAG---ATGGCAGCATTTGTTAAACCAATCTGCACCTCTGCTGTAC 118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 953 LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr 971
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 AGAAGCAAAACTCAG---TATGAGCTGTTCAAAGTGAAGTGTATGATTCGAAGTGG 175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 972 TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla 991
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 TATAAGCTTCGGCTTGGTGTTCACAGACAGAGATATCATCTTCAAAGTAGGGCT 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 992 AlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuAlaAla 1011
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 GCAGTAACTTCAAAAGCTTTT---TGTAGAATGCTCAAGAAACTG-----283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnAlaAlaSerTyrIleGln 1031
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 284 -----GAAACACAGAAATGTCTGCTCCCTACGGATTACG 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1032 -----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg 1049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 TTCTTCTTCAGATGGCTGTGTAT-----CGGAGAAGATTGTTTCAG 358
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1050 GlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGlu 1069
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 CAGAAAGAGCTGCTATCATCTTACAGCATTTATTTAGGACG---TGG-----CAA 406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnGlnArgTrpArg 1089
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 ACCAGAAAACAGTTTTTACTATATAGAAAACGACGATGTTTACAAAATCATCACA 466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1090 AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu 1109
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 467 GCATTTCTGTCTGCAAAAACATCAAGACAAGTCATTTTACAGATCAGACAGTGTTATC 526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyrTyr 1129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 527 ATTATTCAAGCTAGAAAGTAAAGGATTTATACAGAAACGG-----565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1130 SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg 1149
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 565 -----565
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1150 GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln-----1166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 -----AAGTTTCAGGAAATTAATAATAGCACCATATAAATTCAGGCTATCGG 613
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1167 LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg 1186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 AGGAGATATAGAGCCCAAGAAA-----TATTTATGTAAAGTG 649
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 LysCysIleIleAsnValGln-----ArgArgTyrArgAlaThrLeuGlnMetArg 1203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 650 AAAGCTGCCTGCAAGATTCAAGCCTGGTATAGATGTTGGAGACA-----694
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1204 ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPhe 1223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 695 -----CACAAAGAATATCTAGCTGTATTAAGCTGTATAAATTTATCAAGGTTGCTTC 748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1224 ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLysLysValThr 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 749 TATACCAAACTAGAG-----AGAACACCGGTTTTTGAATGTGACAGCATCAGCA 796
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1244 LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnGluTyr 1263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 ATTATCATTCAGAGAAATAGGAGCTATATCTTCCTGCAAGATAGCTCATGACACTTC 856
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1264 LeuHisLeuArgGluValThrIleLysLeuGlnArgPheHisAlaGlnLysSerMet 1283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TTAATGATAAAGA-----871
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1284 ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 872 -----CATCGAGCTGCTTGTGTTGATCCAGACACAT 901
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1304 TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla 1323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 902 TATAGAGGA-----910
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1324 AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer 1343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 -----TATAAGGAAGGACGAGCTCTCTCTCGGCAG-----940
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1344 TyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla 1363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 941 -----AAATCTGCTGTTGATCATCAAAATATATACGAGCAGGAGGCT 988
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1364 MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg 1383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 989 GGAAGCATGAAAGGATAAATATATTGAATTTAAATAATCTACAGTTATCTTACACAGCA 1048
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



1384 ArgTyrArgGlnLeuMetArgGlnLeuValTyrGlnLysGlnArgGln 1403  
1049 CTGGTGGGCTGGCTAGTACGAAAGATTTTAGAA 1093  
1404 AlaValIleArgLeuGlnLysTyrTrpArgSerIleArgAspMetArgLeuCysLysAla 1423  
1094 GCCAAATTCGACTT 1126  
1424 GlyTyrArgGlnLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal 1443  
1127 GCATATTATCACCTGAAT 1144  
1444 GlnAlaArgGlnArgGlnIlePheLeuSerThrIleArgLysValArgLeuMetGln 1463  
1144 1144  
1464 AlaPheIleArgAlaThrLeuLeuMetArgGlnArgGlnPheGluMetLysArg 1483  
1144 1144  
1484 ArgAlaAlaValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg 1503  
1145 -----GCTGTTAGAAATTCAGAGCCTATAAACTTTTACCTGGCTGTGAGAAATGCTAAC 1198  
1504 GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn 1523  
1199 AAGCAGGTT-----AATTGATCATCTGATTACAGAGAGGTTTCAGAGCAGA 1246  
1524 ArgSerMetLysGlnAlaArgGlnGlnPheValGlnLeuArgThrIleAlaValHisLeu 1543  
1247 TTCAAGAAAGAGATTTATTCAGAAATATCATAGCATC----- 1285  
1544 GlnGlnLysPheArgGlyLysArgLeuMetIleGlnArgAsnCysPheGlnLeuLeu 1563  
1286 -----AAAAAGATTCAGCATGAAGTCAAGAAATGTCAGACCGCAGA 1327  
1564 ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe 1583  
1328 AATAGGCTGCATCATGATATACAGAAAGCAGTCCGCAATTTCTCCGTAAGAGCAG 1387  
1584 GlnAlaLeuMetThrProGluMetMetAspLeuIleArgGlnLysArgAlaAlaLysVal 1603  
1388 GAAAAATTCACCTAGT-----GGAATCATTTAAA----- 1414  
1604 IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeu 1623  
1415 ATTCAGGCATTTAGAGAGGCTATTTCTGAGAGAGAAATGATTGTACAAAATTTAAA 1474  
1624 AspIleArgLysArgIleAlaGlnLeuArgGlnGlnAlaLysAlaValAsnSerValArg 1643  
1475 GCTATACGACTAAGTCTTCAAGTTGTTAATAGGAGATTCGAGAGAAACAACTCTAC 1534  
1644 CysLysValGlnGlnAlaValArgPheLeuArgGlyArgPheIleAlaSerAspAlaLeu 1663  
1535 AAAAGAACTGCATTCGCACTTCACTTACCTTTTGACATATAAGCACCTTTCTGCCATTCCT 1594  
1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuMetTrpCys 1683  
1595 GAGGCTTTAAACACCTAGAGGTAGTACTAGA----- 1627  
1684 SerGluPheMetSerThrPheCysTyrGlyIleMetAlaGlnAlaIleArgSerGluVal 1703  
1628 -----TTGCTCCACTTGTGTGAGACATAGCCAGAGTGGCAATTTCTAAA 1678  
1704 AspLysGlnLeuIleGluArgCysSerArg----- 1713  
1679 ATATTGTTTTCATCCGAAGTTGTAATCGCAGTATTCCTTGTATGGAAGTCATCATATAT 1738  
1714 -----IleIleLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730  
1739 GCTGTGCAAGTCTTGTCTTAATGATATTAAGTATGAGAAACTACTTCAGCAGGTTTATGAT 1798

1731 GluGlyGlyLeuValThrIleAlaGlnMetLeuLeu----- 1742  
1799 GTAGAAAATTGATAGATATCTATTGGAGCTTTTGAGATATACCGAGAAAAGCCTGGT 1858  
1743 ---ArgTrpCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVal 1761  
1859 AATAAAGTTGCAGACAAAGCGGAGCAATTTTACAAAACCTTGTGTTTGGCTATT 1918  
1762 PheAlaHisCysProLysLysArgLysIleIleHisAspTyrMetThrAsnProGluAla 1781  
1919 TTACTG-----AAGACAACAATAGAGCCTCTGATGTACGAGTAGTCCAAAGTT 1969  
1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801  
1970 GTTGACCGTATTACAGTCTCTACAACTTACAGCTCATATAACATAAAATG-----AAT 2023  
1802 AlaArgLysProProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817  
2024 ACTGAAAGAATA-----CTTTACAGCAAAAGAAAGATTCTTCTATA 2065  
1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830  
2066 AGCAATTCCTTTTATCCAGAAACACCTGTAAAGCAGCAGAAATAGTTTCAAGACTTAAGCCA 2125  
1831 AspPheGlyIleIleArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850  
2126 GATGGGTTTGAAGAGATACATGAGGAAGAAATCACAATCCCTCGAAGCTATTCAA 2185  
1851 ThrIleLeuCysLysLeuGlnIle 1858  
2186 ATGGTGTGATGATGCTTGGCATT 2209

## RESULT 9

US-10-188-832-81  
; Sequence 81, Application US/10188832  
; Publication No. US20040076955A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions  
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder  
; TITLE OF INVENTION: Cancer  
; FILE REFERENCE: 018501-002330US  
; CURRENT APPLICATION NUMBER: US/10/188,832  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US 60/302,814  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/310,099  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/343,705  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/372,246  
; PRIOR FILING DATE: 2002-04-12  
; NUMBER OF SEQ ID NOS: 207  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 81  
; LENGTH: 2414  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-188-832-81

## Alignment Scores:

Pred. No.: 1,32e-32 Length: 2414  
Score: 437.00 Matches: 210  
Percent Similarity: 35.74% Conservatives: 153  
Best Local Similarity: 21.28% Mismatches: 331  
Query Match: 4.59% Indels: 294  
DB: 17 Gaps: 33

US-09-914-698-1 (1-1861) x US-10-188-832-81 (1-2414)

Qy	913	ArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValLys	932
Db	2	AAATCTGTACGAACATTTACGGCTGCTTTTAGAGGCATGAAGATTGACAAATAATTCGAAA	61
Qy	933	LeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgTyr	952
Db	62	AATGTATCAGAGAAAG---ATGGCAGCCATTGTTAAACCAATCGCATCTGCTGTTC	118
Qy	953	LeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIle---IleThrIleGlnArgTyr	971
Db	119	AGAAGCAAACTCAG---TATGAAGCTGTCAAAAGTGAAGGTGTATGATTCAAGACTGG	175
Qy	972	TrpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla	991
Db	176	TATAAAGCTTCGCGCTTGCTGTTCACAGGAACAGAGTATCATTTCTCAAGTAGGCGCT	235
Qy	992	AlaIlePheLeuGlnArgIleTrpArgArgArgLeuPheAlaLysLysLeuLeuAla	1011
Db	236	GCAGTAACAATTCAAAAGCTTTT---TGTAAGATGGTCAAGAAACTG-----	283
Qy	1012	AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnAlaAlaSerTyrIleGln	1031
Db	284	-----GAAACACAGAAATGTGCTGCCCTACGGATTCAG	316
Qy	1032	-----MetGlnTrpArgThrTyrGlnLeuGlyArgIleGlnArgHisGluPheLeuArg	1049
Db	317	TTCTTCCTTCAGATGGCTGTGTAT-----CGAGAAAGATTTGTTTCAG	358
Qy	1050	GlnArgAspLeuIleMetPheValGlnArgArgMetArgSerLysTrpSerMetLeuGln	1069
Db	359	CAGAAAGAGCTGCTATCACCTTTACAGCATTTATTTAGGACG---TGG-----CAA	406
Qy	1070	GlnArgLysGluPheGlnGlnLeuLysArgAlaAlaIleAsnIleGlnArgTrpArg	1089
Db	407	ACCAGAAAACAGTTTTACTATATAGAAAAGCAGCAGTGGTTTACAAAATCACTACAGA	466
Qy	1090	AlaLysLeuSerMetArgLysCysAsnAlaAspTyrLeuAlaLeuArgSerSerValLeu	1109
Db	467	GCATTCUGTCTGCAAAACATCAAGACAGCTATTTTACAGATCAGACGAGTGTATC	526
Qy	1110	LysValGlnAlaTyrArgLysAlaThrIleGlnMetArgIleAspArgAsnHisTyr	1129
Db	527	ATTATTCAGCTAGAGTAAGTAAGGATTATATCAGAAACGG-----	565
Qy	1130	SerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArgAlaIleMetLysMetArg	1149
Db	565	-----	565
Qy	1150	GluGlnArgGluAsnTyrLeuArgLeuArgAsnAlaSerIleLeuValGln-----	1166
Db	566	-----AAGTTTCAGGAAATATAAAATAGACCATAFAAAATTCAGGCTATGTGG	613
Qy	1167	LysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnAlaTyrLeuArgThrArg	1186
Db	614	AGGAGATATAGAGCAAGAAA-----TATTTATGTAAGAATG	649
Qy	1187	LysCysIleIleAsnValGln-----ArgArgTrpArgAlaThrLeuGlnMetArg	1203
Db	650	AAAGCTGCCTGCAGATTCAGCCTGGTATAGATGTTGGAGAGCA-----	694
Qy	1204	ArgGluArgLysAsnTyrLeuHisLeuGlnThrThrLysArgIleGlnIleLysPhe	1223
Db	695	-----CACAAAGAATATCTAGCTGTATTAAGCTGTAAATAATTTCAAGGTTGCTTC	748
Qy	1224	ArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeuLysLysValThr	1243
Db	749	TATACCAACTAGAG-----AGAACACGGTTTTTGAATGTGAGAGCATTCAGCA	796
Qy	1244	LeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGluArgGlnIleTyr	1263
Db	797	ATTATCAITTCAGAGAAATGAGAGCTATATCTCTCGCAAGATAGCTCATGAACACTTC	856

Qy	1264	LeuHi1sleuAArgGluValThrIleLysLeuGlnAArgPheHisalaGlnLysSerMet	1287
Db	857	TTAATCATAAAAAGA	871
Qy	1284	ArgPheMetArgAlaLysTyrArgGlyThrGlnAlaAlaValSerCysLeuGlnMetHis	1303
Db	872	-----CATCGAGCTGCTGTTTGATCCCAAGCACAT	901
Qy	1304	TrpArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheLeuGlnLeuArgGlnAla	1323
Db	902	TATAGAGGA	910
Qy	1324	AlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetIleLysGlnLeuLysSer	1343
Db	911	-----TATAAGGAGCGAGCTCTCTTCGGCAG	940
Qy	1344	TyrAlaGlnLeuLysGlnAlaAlaIleThrIleGlnThrArgTyrArgAlaLysLysAla	1363
Db	941	-----AAATCTGCTGCTTTCATACAAAAATATATACGAGCGACGAGGCT	988
Qy	1364	MetGlnLysGlnValValLeuTyrGlnLysGlnArgGluAlaIleIleLysValGlnArg	1383
Db	989	GGAAAGCATCAAAAGGATAAAATATATGAATTTAAATAATCTACAGTTATCTTACAGCA	1048
Qy	1384	ArgTyrArgGlyAsnLeuGluMetArgLysGlnIleGluValTyrGlnLysGlnArgGln	1403
Db	1049	CTGCTGCTGCTGGCTAGTACGAAAAAGATTTTAGAA	1093
Qy	1404	AlaValIleArgIleGlnLysTyrTrpArgSerIleArgAspMetArgLeuCysLysAla	1423
Db	1094	GCCAAAATTCGACTT	1126
Qy	1424	GlyTyrArgArgIleArgLeuSerSerLeuSerIleGlnArgLysTrpArgAlaThrVal	1443
Db	1127	GCATATTATCACCTGAAT	1144
Qy	1444	GlnAlaArgGlnArgGluIlePheLeuSerThrIleArgLysValArgLeuMetGln	1463
Db	1144	-----	1144
Qy	1464	AlaPheIleArgAlaThrLeuLeuMetArgGlnGlnArgGluPheGluMetLysArg	1483
Db	1144	-----	1144
Qy	1484	ArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaMetLeuLysAlaArg	1503
Db	1145	-----GCTGTAGAAATTCAAAGAGCCTATAAACTTTACCTGGCTGGAAGAATGCTAAC	1198
Qy	1504	GlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgLysPheArgAlaAsn	1523
Db	1199	AAGCAGGTT-----AATTCACTCATCTGTATTACAGATGGTTTCGAGCAAGA	1246
Qy	1524	ArgSerMetLysGlnAlaArgGlnGluPheValGlnLeuArgThrIleAlaValHisLeu	1543
Db	1247	TTACAGAAAGAGATTTATTCAGAAATATCATAGCATC	1285
Qy	1544	GlnGlnLysPheArgGlyLysArgLeuMetIleGlnGlnArgAsnCysPheGlnLeuLeu	1563
Db	1286	-----AAAAAGATTGAGCATCAAGGTCAGAAATGCTCTGAGCCACGCA	1327
Qy	1564	ArgCysSerMetProGlyPheGlnAlaArgAlaArgGlyPheMetAlaArgLysArgPhe	1583
Db	1328	AATAGGCGTCGATCAGTAATACAGAAAGCAGTGGCCCATTTCTCTCCCGTAAAGACAG	1387
Qy	1584	GlnAlaLeuMetThrProGluMetMetAspIleIleArgGlnLysArgAlaAlaLysVal	1603
Db	1398	GAAAAATTCACACTAGT-----GGAATCATATAA	1414
Qy	1604	IleGlnArgTyrTrpArgGlyTyrLeuIleArgArgGlnLysHisGlnGlyLeuLeu	1623
Db	1415	ATTTCAGGCATTATGAGAGGCTATCTCTGGAGGAGAGAAAATGATGTACAAAATAAAA	1474
Qy	1624	AspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaValAsnSerValArg	1643

```

Db 1475 GCTATACGACTAAGTCTTCAAGTTGTTAATAGGAGATTTCGAGAGAAAACAACTCTAC 1534
QY CysLysValGlnGluAlaValArgPheLeuArgGlyArgPheLeuAlaSerAspAlaLeu 1663
Db 1535 AAAAGAACTGCACTTGCCTTACCTTGTGATATAGACACCTTCTGCTGCTTCTT 1594
QY 1664 AlaValLeuSerGlnLeuAspArgLeuSerArgThrValProHisLeuLeuMetTrpCys 1663
Db 1595 GAGGCTTAAACACCTAGAGGTAGTACTAGA----- 1627
QY 1684 SerGluPheMetSerThrPheCysTyrGlyLeuMetAlaGlnAlaLeuArgSerGluVal 1703
Db 1628 -----TTGTCTCCACTTGTGTGAGAACATGCGCCAGAGTGAGGCAATTTCTAAA 1678
QY 1704 AspLysGlnLeuLeuGluArgCysSerArg----- 1713
Db 1679 ATATTGTTTGAATCCGAGTTGTAATCGCAGTATCTTGTATGGAAGTCATCAGATAT 1738
QY 1714 -----IleLeuLeuAsnLeuAlaArgTyrAsnSerThrThrValAsnThrPheGln 1730
Db 1739 GCTGTGCAAGTCTTGTCTTAATGATCTAAGTATGAGAAACTACTTCAGCAGTTTATGAT 1798
QY 1731 GluGlyCysLeuValThrIleAlaGlnMetLeu----- 1742
Db 1799 GTAGAAAATTTGTATAGATATCTATTGGAGCTTTTGCAGATATACCGAGAAAAGCCTGGT 1858
QY 1743 ---ArgTrpCysAspLysAspSerGluLeuPheAsnThrLeuCysThrLeuIleTrpVal 1761
Db 1859 AATATAGTTCCAGACAAAGCGGAGCATTTTACAAAAAATTGTTGTTGGCTATT 1918
QY 1762 PheAlaHisCysProLysLysArgLysIleLeuHisAspTyrMetThrAsnProGluAla 1781
Db 1919 TTACTG-----AAGACAAATAAGAGCCTCTGATGACGAAGTAGTCCAAAGTT 1969
QY 1782 IleTyrMetValArgGluThrLysLysLeuValAlaArgLysGluLysMetLysGlnAsn 1801
Db 1970 GTTGACCGTATTACAGTCTCTCAAACTTACAGCTCATAAACATAAAATG-----AAT 2023
QY 1802 AlaArgLysProProMetThrSerGlyArgTyrLysSerGlnLys----- 1817
Db 2024 ACTGAAGAATA-----CTTTACAAAGCAAAAGAAATCTTCTAT 2065
QY 1818 ---IleAsnPheThrProCysSer-----LeuProSerLeuGluPro 1830
Db 2066 AGCATTCCTTTTATCCAGAAACACCTGTGAGGACGAGATAGTTTCAAGACTTAAGCCA 2125
QY 1831 AspPheCysIleLeuArgTyrSerProTyrThrPheIleSerSerValTyrAlaPheAsp 1850
Db 2126 GATTGGGTTTGTGAAGAGAGATAACATGGAAGAAATCACAAATCCCTGTCAAGCTATTCAA 2185
QY 1851 ThrIleLeuCysLysLeuGlnIle 1858
Db 2186 ATGGTGATGATACGCTTGGCAIT 2209

RESULT 10
US-10-437-963-53524/c
; Sequence 53524, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

```

```

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 53524
; LENGTH: 3069
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55716C.1
US-10-437-963-53524

Alignment Scores:
Pred. NO.: 1.34e-23 Length: 3069
Score: 348.50 Matches: 239
Percent Similarity: 36.04% Conservative: 187
Best Local Similarity: 20.22% Mismatches: 440
Query Match: 3.66% Indels: 317
DB: 1.7 Gaps: 42

US-09-914-698-1 (1-1861) x US-10-437-963-53524 (1-3069)

QY 419 AlaIleSerProLysLysGlnArgValGluAsnThrThrLeuProArgSerAlaAla 438
Db 3051 GCGGTCCTCTCCCTCCCTTCCGCGAC---CTTCCAACTCCGACCCCGAACCCGAGCGC 2995
QY 439 ProAlaAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLys 458
Db 2994 CCGCTAACCCCAATCTCGGCTCCAGAGGAACCCCTCCCT----- 2950
QY 459 LysPheLysLeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysVal 478
Db 2949 -----TCGCGCACCCCGACGACCCGG 2929
QY 479 ArgAspThrSerIleGlnProSerValLysLeuTyrAspSerGluLeuTyrMetGlnThr 498
Db 2928 CGCGCGCGGGGCTCTCTCTCTCT----- 2905
QY 499 CysIleAsnProAspPropheAlaAlaThrThrIleAspPropheLeuAlaSerThr 518
Db 2904 -----CCTCGCCCGCGCGGACCGCGACCGCCCTCGCCCGCCCTCGCGCGC 2854
QY 519 MetTyrLeuAspGluGln-----AlaVal 526
Db 2853 CTGAGCTTCGACAGTCGCGATCGCGCGCGCGCGAGTCGCGCGCGCGCGCGCTC 2794
QY 527 AspArgHisGlnAlaAspPheLysLysTyrLeuAsnAlaLeuValSerIleProAla--- 545
Db 2793 CGCGCTTCTGCTCTCTCTCGCCACGTCGTGCTCTCTCTCGCGGACCGCTCGGCC 2734
QY 546 -----AspLeuAspAla---Asp 550
Db 2733 TCGCGCTGCGCGCTTCCGCGCGCGCGCGAGGTACCGCGAGCTCTCTCTCAGCGC 2674
QY 551 LeuAsnAsnLysIleAspVal-----GlyLysLeuPheAsnGlu 563
Db 2673 GTGAGGCAAGCGCGACGCGGTGATGGGAGCGCGCGGTGCGGAGCGCGCGAGCGC 2614
QY 564 ValArgAsnLysGluLeuValAlaProThrLysGluGlnSerMetAsnTyrLeu 583
Db 2613 CACCGGGGTGTGAGGATCGCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 2554
QY 584 ThrLysTyrArgLeu---GluThrLeuArgLysAla-----AlaValGlu 597
Db 2553 ATGGCGGCTTCTGTGAGGATTCCTTGAGGAGTTCCTTGAGGAGTGTGAGGAGTGTGAG 2494
QY 598 LeuPhePheSerGluGlnMetArgLeuProCysSerLysValAlaValTyr----- 614
Db 2493 AGGATGGGCAAGCAGATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2434
QY 615 -----ValAsnLysGlnAlaLeuArgIleArgSerAspArgAsnLeuHisLeu 630
Db 2433 ATCTGCAAAACATTGAGGAGTTAAAGATGAAAGAACATTGCCCTTGTCTTCT 2374
QY 631 AspValValMetGlnArgThrIleLeuGluLeuLeuCysPheAsnProLeuTrpLeu 650

```

2373	Db	GATCTCAGGCTGAGAGATAAGCAATAAGAACTCTTCATGTGTGTTATCAATCTCTAAGTGGCTC	2314
651	Qy	ArgLeuGlyLeuGluValValPheGly-----GluIysIleGlnMetGln	665
2313	Db	AGGATAGACTGCATATTGTCCTCGGGGTGATCTCTGGATACAGAATGAATCGCAGAAA	2254
666	Qy	SerAsnArgAspIleValGlyLeuSerThrPheIleLeuAsnArgLeuPhe-----	682
2253	Db	AAGGACAAAGAAAGTTGCTTCTCAATTTGTTCTCGAGAAAGCAATGTTTGTTCATATG	2194
683	Qy	-----ArgAsnIysCysGluGluGlnArgTyrSerLysAlaTyrThr	696
2193	Db	AATACACACCTTCCTCGCTCTCAACAGCGCCCGAGGGGCACCGTAGAGCATGT-----	2140
697	Qy	LeuThrGluGluTyrAlaGluThrIleLeuIysHisSerLeuGlnIysIleLeuPheLeu	716
2139	Db	-----TATGCTGAAGCGCCAGCAACAGATATCTGAAAGAGTTATTTCTTCCT	2092
717	Qy	LeuProPheLeuAspGlnAlaLysGlnLysArg-----	727
2091	Db	GTTCCTGCCTTAGACAGAGCTAAATAAGAAAGTGTTTACCTTCGGAATCTGGAATCGCAT	2032
728	Qy	---IleValIysHisAsnProCysLeuPheValIysLysSerProHisLysGluThrLys	746
2031	Db	GGCTCGAGTGTGATCCCTTTGTTATTTGTCCGACGCGAAATAAAGTCTAGTCGG	1972
747	Qy	AspIleLeuLeuArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGlu	766
1971	Db	CAAAATGTTCAGGAGTCATTGGGGGAAGTCATGTCATGAGAGGTGATCTATTGATGCAT	1912
767	Qy	LeuArgArgLeuGlyTyrValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyr	786
1911	Db	CTCAATATCATGGGATATAAATTGAATTATCAGCAGCTGCTCTATCAGAGTAGTATGATTC	1852
787	Qy	AlaPheAsnAsnLeuAlaValAspIleuArgAspGlyValargIleuThrArgValGlu	806
1851	Db	ACTGTAGGAAATTTATTTGAAGATCTTCAAGATGGCAATCTCTGTAGAAATTTATTCAG	1792
807	Qy	ValIleLeuLeuArgAspAspLeuThrArgGlnLeuArgVal-----ProAlaIleSer	824
1791	Db	-----CTGCCTAACATCTGATGCATCGATTTATTTGAAGTGATGCTCCATCAGATACA	1738
825	Qy	ArgLeuGlnArgIlePheAsnValIysLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPhe	844
1737	Db	TATAAGAAAAAGATTGCTATAACTGCACCATGGCAATCCAATATATCAACGAGCTGGGTTT	1678
845	Qy	GlnLeu-----GlyClyAspIleAlaAlaGlnAspIleValaspGlyHisArg	860
1677	Db	CCGTTTGTCTGATGCAGATGGACTCTCAATCTCTCGGAAGACATGTGCAATGGGGATAAG	1618
861	Qy	GluLysThrLeuSerLeuLeuTrpGlnLeuIleTyrLysPheArgSerProLysPheHis	880
1617	Db	GAGCTAATTCAGCGTTGCTTTGGACACATGTTTATATATGTCAGTTACCAAGTTAGTA	1558
881	Qy	AlaAlaAlaThrValLeuGlnLysTrpTrpArgHisTrpLeuHisValValIleGln	900
1557	Db	AATGAAACTTCAGTAGCTCAGGAA-----	1534
901	Qy	ArgArgIleArgHisLysGluLeuMetArgHisArgAlaAlaThrValIleGlnAla	920
1533	Db	-----ATATCAAGGCTAAAGGCACCGCATGATCAGAACATCA	1498
921	Qy	ValPheArgGlyHisGlnMetArgLysTyrValIysLeuPheLysThrGluArgThrGln	940
1497	Db	ATC-----TCTGAAATGCAAGTCACAG	1477
941	Qy	AlaAlaIleIle-LeuGlnLysPheThrArgTyrLeu-AlaGlnLysGlnLeuTyrG	960
1476	Db	ACTGGTTTGCTTTATGATTGGATTTCAGGAAACATTTATCTGATTTGTGAAAGCAATGTC	1417
960	Qy	InSer-----TyrHisSerIleIleThrIleGlnArgTrp	972
1416	Db	AGTTGTCTATAAACACAGACATGATTGCTGATATCACAC-----TTATCAATTTACACATT	1360

972 rpArgAlaGlnGlnLeuGlyArgGlnHisArgGlnArgPheValGluLeuArgGluAla- 991  
||:||||:|  
1359 GGAAAAAGT-----GCTT 1348

992 AlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaIleLysLysLeuAlaLa 1011  
|||:|||||:|  
1347 GCACAAATTCCTCAGGACCTACCTGGTTGGAATATCTAGCTAATGATGATATTTGTT 1288  
||:||||:|

1012 AlaGluThrAlaArgLeuGlnArgSerGlnLysGlnGlnAlaAlaSerTyrlleGln 1031  
||:||||:|  
1287 GAAAAAAGTGA-----ATAATCTG 1267

1032 MetGlnTrpArgThrTyrlleGlnGlyArgIleGlnArgHisGluPheLeuArg- 1049  
||:||||:|  
1266 CTCGCAATTCCTCATCTGCATTAATGTCCAGACGATGGAGCACTAAGAATCTG 1207  
||:||||:|

1050 -----GlnArgAspLeuIleMetPheValGlnArgArgMetArgSer 1063  
||:||||:|  
1206 ATCGATTCAAAGTTGGATCATCAAAGTCTGGTTACTGAGGTTTCTCTAGCGGTAGATCT 1147  
||:||||:|

1064 LysTrpSerMetLeuGlnArgLysGluPheGlnGlnLeuLysArg----- 1079  
||:||||:|  
1146 CGA---GGAAACAACGTACGAAGTGCCATTTCCTCAGACAGGAGACAGATGGTTCA 1090  
||:||||:|

1080 -----AlaAlaIleAsnIleGlnGlnArgTrpArgAlaLysLeuSer 1093  
|||:|||||:|  
1089 CGCAGTACCCGAGAGTGGCTGCAACAGTTATACAACTCAAGCTAAGAGACTTAATGCA 1030  
|||:|||||:|

1094 MetArgLys-----CysAsnAlaAspTyrlleAla 1103  
|||:|||||:|  
1029 ATGAGCAAAVACTGTAAGCTTAAGAAATGCAACACAGCCATCAACAGGGCGCATGTCCT 970  
|||:|||||:|

1104 LeuArgSerValLeuLysValGlnAlaTyrlleArgLysAlaThrIleGlnMetArgIle 1123  
|||:|||||:|  
969 GCTTCATCCAGTCTCTGAAGAGCAATTGCTGATAGTAGCTGC-----ATA 925  
|||:|||||:|

1124 AspArgAsnHisTyrlleTyrlleSerLeuArgLysAsnValIleCysLeuGlnGlnArgLeuArg 1143  
|||:|||||:|  
924 GATTCAGCCCACTAAGCTTGTGTGTGGAGATGATGTGGACTGCAGCACTAATCTGTCAA 865  
|||:|||||:|

1144 AlaIleMet-----LysMetArgGluGlnArgGluAsnTyrlleArgLeuArgAsnAla 1161  
|||:|||||:|  
864 GTTTTGTGTTCTATCATGATCCAGTATCTACCAAGGTGCATCTCTCTGTTCTCGCAGGAAGCA 805  
|||:|||||:|

1162 SerIleLeu-----ValGlnLysArgTyrlleArgMetArgGlnGlnMetIleGlnAspArg 1179  
|||:|||||:|  
804 ATGGCTGCTCGAAAGATACAGTTTGCCTATAGAAGATTGCCACACAGAAATAGGTTCAAGA 745  
|||:|||||:|

1180 AsnAlaTyrlleArgLysCysIleIleAsnValGlnArgArgTrpArgAlaThr 1199  
|||:|||||:|  
744 ATTTGCG-----GCTGCATTAAGATCCAAAGCCCAATTCGCGCTGT- 706  
|||:|||||:|

1200 LeuGlnMetArgArgGluArgLysAsnTyrlleHisLeuGlnThrThrLysArgIle 1219  
|||:|||||:|  
705 -----TTC 703  
|||:|||||:|

1220 GlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuGlnLeu 1239  
|||:|||||:|  
702 TCTGTGCGGATACGTTTCAAAGGCAAAATT-----CAAAATATTACTACCATCCAAGCT 649  
|||:|||||:|

1240 LysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgLysGlu 1259  
|||:|||||:|  
648 GTAGCGAGGCTTGGTCAACCGGATCTGGGAAGAGATCTGATGTTGTTATTCAGGCTCAT 589  
|||:|||||:|

1260 ArgGlnGluTyrlleHisLeuArgGluValThrIleLysLeuGlnArgArgPheHisAla 1279  
|||:|||||:|  
588 GTCCGGGGATGGATT---GCACGCGAGACTGCTGTGTAGAAACAAGACGCT----- 541  
|||:|||||:|

1280 GlnLysSerMetArgPheMetArgAlaLysTyrlleArgGlyThrGlnAlaAlaValSerCys 1299  
|||:|||||:|  
540 -----ATTAAGTATA 532  
|||:|||||:|



Db 763 CTCAGCTAAGTATGCCATGCGATATCTTTGCAACATGTCAGTGGTCTCTGCCAGTGAGGCCAA 822  
Qy  
Qy 357 yGlnGlnGluPheLeuPheAsnHisSerGluIleuAla----- 370  
Db 823 TGTGGAGAAAGGCTCTTGACCTCAACCCCATCATGGAGTCCANTGGAAATGCTAAAC 882  
Qy 371 -----GlnSerSerArgPheAsnLeuHis-----GluValGly-----Ar 382  
Db 883 AACACGAGATGATATAGCAGCCGTTTGGGAAGTATATGAGATGGTTGGTTGATAAGAG 942  
Qy 382 gLysSerValnysGlySerProValnysAsnProHisLysArgArgSerHisGluLeuSe 402  
Db 943 ATATCGAATCATTGGTGCATATAGCAATCTATCTTTTAGAGAAATCCAGAGTGTATT 1002  
Qy 402 rPheSerAspAlaProSerAsnGluSerLeuTyArg----- 414  
Db 1003 CCAGGCAGAGAGAGAGAAATCATATCTTCTATCAGCTTTGTGCCTCAGCAAGATT 1062  
Qy 415 ----- 420  
Db 1063 ACCTGAATTTAAATGCTACGATAGGAAATGCAGATACTTTAATTACACAAACAAGG 1122  
Qy 420 eSerProProLysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAl 440  
Db 1123 AGGCAGTCTCTGTGATGAAGAGTGGATGATGCAAGAGGAGATGGCACATCTAGGCAGGC 1182  
Qy 440 aAsnAlaSerAlaArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysph 460  
Db 1183 CTGCACCTTGTAGGAATTAGTAATCTCAT-----CAATGGGNAITTT 1227  
Qy 460 eLys-----LeuAlaGlnThrMetSerLeuMetLysLysProAlaThrProArgLysValAr 479  
Db 1228 CCGAATACTTCTGCGCATCTCTACTTAGCAATGTTGATTATACATCCGAGATGCAGA 1287  
Qy 479 gAspThrSerIleGlnProSerValLysLeuTyArgSerGluLeuTyMetGlnThrCy 499  
Db 1288 CAGCTGCACAAATCTCCCAAG-----CATGAACCTCTCTG 1323  
Qy 499 sIleAsnProAspProPheAlaAlaThrThrThrIleAspProPheLeuAlaSerThrWe 519  
Db 1324 CATC----- 1327  
Qy 519 tTyLeuAspGluGlnAlaValAspArgHisGln----- 530  
Db 1328 -TTCTGACCTCATGGTGTGGACTATGAGGAGATGTCTACTGGCTCTGCCATCGGAA 1386  
Qy 531 -----AlaAspPheLysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAs 548  
Db 1387 ACTGGCTACTGCCACAGACATACATCAAGCCCATCTCCAAGCTGCGAGGCCAGCAATGC 1446  
Qy 548 pAlaAspLeuAsnAsnLysIleAspValGlyLysLeuPheAsnGluValArgAsn----- 566  
Db 1447 CCGGATGCTTTGGCCAAAGCACATCTATGCCAAGCTCTTAACTGGATGTGATATAATGT 1506  
Qy 567 -LysGluLeuValValAlaProThrLysGluGlnSerMetAsnTyLeuThrLysTy 586  
Db 1507 CAATCAGCTCTCCATCTGCTGTCAACAGCAGCTCTTTATTTATGGTGTCTAGACATTTA 1566  
Qy 586 rArgLeuGluThrLeuArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLe 606  
Db 1567 CGGATTTGAAACATTTTGAGATAAATAGTTTGAACAGTTT----- 1606  
Qy 606 uProCysSerLysValAlaValTyValAsnLysGlnAlaLeuArgIleArgSerAspAr 626  
Db 1607 ----TGCAATAAT-----TATGCAAAATGAA-----AACTACACCACAATT 1644  
Qy 626 gAsnLeuHisLeu-----AspValValMetGlnArgThrIleLeuGluLe 641  
Db 1645 CAATATGATGTCCTTCAAAATTTGGAGCAAGAATAATATGAAGCAACAAATTCATGGAC 1704  
Qy 641 uLeuLeu-----CysPheAsnProLeuTyTrpLeuArgLeuGlyLe 654  
Db 1705 ACTCATAGATTTTATGATAATACAGCCTTTGTATTAATCTTATAGAATCAAAACTAGGCAT 1764

Qy 654 uGluValValPheGlyGluLysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSe 674  
Db 1765 TCTAGATTTACTGGATGAGGAATGCAAGATGCTTAAAGGCACAGATGACACCTGGGCCCA 1824  
Qy 674 rThrPheIleLeuAsnArgLeuPheArgAsnLysCys-----GluGluGlnArgTy 691  
Db 1825 AAAATTTGACACACACATTTG-----AACAAATGTGCACCTCTTTGAAAAGCCTCGTCT 1878  
Qy 691 rSer---LysAlaTyx----- 695  
Db 1879 ATCAACAAAGACTTTTCATCTCCAACTTTTCTGCACAAAGTGAATACCACTGTGAAG 1938  
Qy 696 -----ThrLeuThrGluGluTyzAlaGluThrIleLysLysHi 708  
Db 1939 ATTTCTCGAAAGAAATAAGACACCGTTTGAAGAAACAAATTAAGATCTTTAAATCAAG 1998  
Qy 708 sSerLeuGlnLysIleLeuPheLeuLeuProPheLeuAspGlnAlaLysGlnLysArgIl 728  
Db 1999 CAAGTTTAAAG-----ATGCTACCAGAACTATTTCAGATGATGATGAGAGGCCAT 2046  
Qy 728 e-----VallyLysHisProCysLeuPh 736  
Db 2047 CAGTCCAACTTCAGCCACTCTCAGGGCGCACACCCCTCACAGAACTCTGCAAGGCC 2106  
Qy 736 eValLysLysSerPro-----HisLysGluThrLysAspIleLeuLe 750  
Db 2107 CACCAAGGCGAGACAGGCCAAATGGCCAAAGAGCACAAAGAAACAGTGGGGCATCAGTT 2166  
Qy 750 uArgPheSerSerGluLeuLeuAlaAsnIleGlyAspIleThrArgGluLeuArgLe 770  
Db 2167 CAGAACTCTCCCTGACCTGCTT-----ATGGAGACACT 2199  
Qy 770 uGlyTyzValLeuGlnHisArgGlnThrPheLeuAspGluPheAspTyzAlaPheAsnAs 790  
Db 2200 CAATGCCACTACCTCTACTATGCGCTGTATCAAGCTTAATCACTTCAAGTTTC---CC 2256  
Qy 790 nLeuAlaValAspLeuArgAspGlyValArgLeuThrArgValValGluValIleLeuLe 810  
Db 2257 ATTCAAGTTTGTAGAAAGAGGCGAGTGCAGCAGCTGAGAGCATGTGGTGTCTCTG--- 2311  
Qy 810 uArgAspAspLeuThrArgGlnLeuArgValProAla----- 822  
Db 2312 -----GAAACCATCCGAATCAGTCGGCCCGGTTTCCCTCAGCGGTGAC 2355  
Qy 823 -----IleSerArgLeuGlnArgIlePheAsnValLysLeuAlaLeuGly--- 837  
Db 2356 TTACCAAGAAATTTTCAGCCCTACCGTGTCTTAATGAAGCAGAAAGATGTCTGAGTGA 2415  
Qy 838 -----AlaLeuGlyGluAlaAsnPhedI 845  
Db 2416 CAGAAAGCAACATGCAAGAAATGTGTAGAGAAACTGATCTGGACAAGACAAATACCA 2475  
Qy 845 nLeuGlyGlyAspIleAlaAlaGlnAspIleValAspGlyHisArgGluLysThrLeuSe 865  
Db 2476 GTTTGGT-----AAGACAAAGATCTTTT 2499  
Qy 865 rLeuLeuTrpGlnLeuIleTyx-----LysPheArgSerProLysPheHisAlaAlaI 883  
Db 2500 CGTCCCGCTCAAGTGGCTTATCTAGAAAAATTTGAGAGCTGACAACTGAGAGCTGCTG 2559  
Qy 883 aThrValLeuGlnLysTrpTrpArgArgHisTrpLeuHisValValIleGlnArgArgIl 903  
Db 2560 CATCCGATCCAGAAAG---ACCATCCGAGGGTGGCTG----- 2593  
Qy 903 eArgHisLysGluLeuMetArgArgHisArgAlaAlaThrValIleGlnAlaValPheAr 923  
Db 2594 -CTGAGAAGAAGTAGTACCTACGCATCGGGAAGGACGCATCACCATGCAGAGATCGTGG 2652  
Qy 923 gGlyHisGlnMetArgLysTyzValLysLeuPheLysThrGluArgThrGlnAlaAlaI 943  
Db 2653 GGGCTACCAAGGCCCGCATGCTATGCTAAGTTTCTGCGC-----AGAACCAGACGCAAC 2706



Db 4687 CAAGAAGGAGGATGAGCAAAACCTGTTAAGAACCTGATTTGGAACTGAGCCACCGTGG 4746  
Qy 1549 YLysArgLeuMetTleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProG 1569  
Db 4747 TGTAGCAGTCANTTGG-----ATTCAGG 4770  
Qy 1569 yPheGlnAlaArgAlaArgGlyPheMetAlaArg-----Ly 1581  
Db 4771 ATTACCGGCATATATCTGTTTCATGTGTGTTTCGACATGCTGACTACCTGATGATCA 4830  
Qy 1581 sArgPheGlnAlaLeuMetThrProGluMetMetAspLeuLeuArgGlnLysArgAlaAl 1601  
Db 4831 GAAAGTAAAGTCGTGCTTAACATCAACAATTAACAGCATC-----AA 4872  
Qy 1601 aLysValIleGlnArg-----TyrTrpArgGlyTyrLe 1612  
Db 4873 AAAAGTATTGAGAAAGAGAGTGATGATTTGAAACCGTCTCTCTGCTCTTAACAC 4932  
Qy 1612 uileArg----- 1614  
Db 4933 ATGCGGATTTTGCACCTGCTTGAACACAGTACAGTGGAGAGGCGCTTTATGAAGCACA 4992  
Qy 1615 ----ArgArgGlnLysHisGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuAr 1633  
Db 4993 CACATCTGCCAGATGAACACTCCCTCACCAATTTT-----GACCTGGCTGATCG 5046  
Qy 1633 gGlnGluAlaLysAlaValAsnSerValArgCysLysVal--GlnGluAlaValArgPh 1652  
Db 5047 GCAG-----GTGCTGAGTGTGCTGGCCATTCAGATCTACCAGCAGCTGTCGCGGT 5097  
Qy 1652 eLeuArgGly-----ArgPheIleAlaSerAspAlaLeu----- 1663  
Db 5098 GTTAGAGAACATCCTTCAGCCCAATGATGTTCTCAGGATGCTGGAACATGAACGATTC 5157  
Qy 1664 -----AlaValLeuSerGlnLeuAspAr 1671  
Db 5158 GGGCGTGTCTGGGTGAAGCCCAAGGTTTGAGAAAGCGAACCTCCAGTATCGCGCATGA 5217  
Qy 1671 gLeuSerArgThrValProHisLeuLeuMetTyrCysSerGluPheMetSerThrPheCy 1691  
Db 5218 GGGCACTACACATGACATCCATCTCCGCGAGCTCACTCTCCACTCGGTCAATGTG 5277  
Qy 1691 sTyrGlyIleMetAlaGlnAlaIleArgSerGluValAspLysGlnLeuLeuArgCy 1711  
Db 5278 TCAGCATGGCATGACCTGAACTGATCAAGCAGGTGTGCAAGCAGATGTTCTACATCAT 5337  
Qy 1711 sSerArgIleIleLeuAsn----- 1717  
Db 5338 AGGGCCATCACCTGGAACACCTTCTCTCGGAAGACATGTGCTCTCGAGTAAAGG 5397  
Qy 1718 ----LeuAlaArgTyrAsnSerThrThrVal-----AsnThrPh 1729  
Db 5398 CATCGATCAGGTACATGTCAGTCACTCGAAGATGGCTGGTGCAGCAAGATCTGAT 5457  
Qy 1729 eGlnGluGlyGly-----LeuValThrIleAlaGlnMetLeu----- 1741  
Db 5458 GAATAGTGGGCTAAAGAAACCTCGAACCTCTCATTCAGGCTGCTCAACTTTTTCAGT 5517  
Qy 1742 -LeuArgTyrCysAspLysAspSerGluIlePheAsnThrLeuCysThrLeuIleTrpVa 1761  
Db 5518 GAAAGAGAAACAGATGATGATGAGAA-----GCCATTTGTTCTAT-GTGCAATG 5567  
Qy 1761 lPheAlaHisCysPro----- 1766  
Db 5568 CTTTAACTACTGCCAGATGTTGAAAGTGTGAATTGTTATCTCCAGTTAATGAGTTG 5627  
Qy 1767 ----LysLysArgLysIleIleHisAspTyrMetThrAsn 1778  
Db 5628 AAGAAAGAGTCTCTGTGTCGTTCATTCGTACTATACAGAT 5667

RESULT 12

US-09-954-456-1601

; Sequence 1601, Application US/09954456

; Patent No. US20020115057A1

; GENERAL INFORMATION:

; APPLICANT: Young, Paul

; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can

; TITLE OF INVENTION: Sets

; FILE REFERENCE: 689290-76

; CURRENT APPLICATION NUMBER: US/09/954,456

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: US/60/233,617

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: US/60/234,052

; PRIOR FILING DATE: 2000-09-20

; PRIOR APPLICATION NUMBER: US/60/234,923

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,134

; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,637

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,638

; PRIOR FILING DATE: 2000-09-26

; PRIOR APPLICATION NUMBER: US/60/235,711

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,720

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,840

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: US/60/235,863

; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 2276

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1601

; LENGTH: 14800

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-954-456-1601

Alignment Scores:

Pred. No.:	7.8e-20	Length:	14800
Score:	322.00	Matches:	363
Percent Similarity:	35.69%	Conservative:	289
Best Local Similarity:	19.87%	Mismatches:	682
Query Match:	3.38%	Indels:	493
DB:	9	Gaps:	72

US-09-914-698-1 (1-1861) x US-09-954-456-1601 (1-14800)

Qy	77	LysLysArgMetSerAlaAlaAlaProProSerLysGlnThr-----Trp	92
Db	3062	AGGAAGAGTCTCGCTGCAGCGCTGCATCTCCGAGCTCAAAGACATCCGGCTGCAGCTGG	3121
Qy	93	ArgValThrAlaProSerArgProAlaAlaTrpAla-HisProProGlnAla-ProL	112
Db	3122	-----AGCCTGTGAGAGCGCGACCGTGCACCGCTCGGGTGGCGC	3163
Qy	112	euValGluLysAsnValTyrIlystThrProGlnGluProValTyrIleSerProGlnP	132
Db	3164	TGGACAAAGAGCGCGCAGCGAGTGTGCCAGCGC-----ATGCCGAGCAGC	3211
Qy	132	roArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuLeuAspValIleA	152
Db	3212	AGAAGGCACAGGCAGAG-----	3228
Qy	152	spAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePhe-	171
Db	3229	-----GTGAGGGGGCTGGGCAGAGGGGTGCCCGGCTCTCTG	3265
Qy	172	-----ProAspAsnLeuAlaAlaTrpPro-----ThrProThrLeuLysG	185
Db	3266	CCGAGGCGGAGAGGTCTTGGCCCTACAGAGCATCGCTTCGCGGCCGCCACGCTCGCT	3325
Qy	185	lyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuG	205
Db	3326	CGGAGCTGGAG-----CTGACGCTGGGCAAGCTGG	3355



205 luaspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnIleS 225  
 3356 AGCAGGTCCGCGAGCTCTGCCATCTAC---CTGGAGAGAGCTCAAGACCATCAGCCTGG 3412  
 225 erLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysT 245  
 3413 TGTATCCGCGGCACGACGAGGGCCGAGGAGGTGCTCAGGGCCCAACGAGGAGCAGCTCAAGG 3472  
 245 hrThrThrIleValHisAlaThrHisThrArg-----AlaLeuAlaCysIleHisG 262  
 3473 AGGCCACAGGCGCGCGCCGACCTCCCGAGGCTCGAGGCGCACCAAGGCGCTCTCTGAAGA 3532  
 262 luGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeuLysA 282  
 3533 AGCTGGCGGGCCCGAGCGACAGCAGCCACGCTTCGACGCGCTCGCGGATGAGCTGC 3592  
 282 rgAspIleLysLeuValGlySerProLeu---ArgIlySerGluSerMetLysAspL 301  
 3593 GGGGGGCACAGGAGGTGGCGAGCGACTCGCAGCAGCGCGACGGGAGCGGGACTTGGAGG 3652  
 301 euSerLeuLeuSerProGlnThrLysThrAlaIleGlnGlySerMetProAsnLeuAsnG 321  
 3653 TGGAGCGCTCGCGGAGCGGGTGCCTCCAGGTGCTTGGAGCGCTGCGCAGGCTGTGCTGCC 3712  
 321 luMetLysIleArgSerIleGlu-----GlnAsnArgTyrTyrGlnG 335  
 3713 AGACCGACTTGGCGCAGCGGAGCTCGAGCACTGGGCGCCCGCAGCTCGCTTACTACCGCG 3772  
 335 lu-----GlnIleGlnIleL 341  
 3773 AGAGTCAGACCCCTTGGGGCGCTGGTGTCAGCAGCGCGCAGCGCGCGCAGAGCAGATCC 3832  
 341 ysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGln---- 359  
 3833 AGGCCATGCCGTGGCGCGACAGCAGCGAGCTGTGGCGGAGCAGCTGCGGCGAGCAGCGGCC 3892  
 360 ----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe----- 375  
 3893 TGTGTGAGGAGATCGAGCGCCAGCGGAGAGGTGAGGAGTCCGAGAGGTTCGCGAAGC 3952  
 376 ----AsnLeuHisGluValGlyArgLysSerValLysG 387  
 3953 AGTACATCAACGCCATCAAGGACTATGAACTCCAGCTGGTGAGCTACAGGCG---CAGC 4009  
 387 lySerProValLysAsnProHisLysArg----- 396  
 4010 TTGAGCGGTGGCTCCCGCGCAAGAAGCCCAAGGTCCAGTCGGGATCAGAGAGTGTC 4069  
 397 ----ArgSerHisGluLeuSerPheSerAspAlaProSer--- 408  
 4070 TCCAGGATACGTGGACCTTGGTACGCTACGACGAGCTGACACACTGACGAGCGCAGT 4129  
 409 ----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProL 424  
 4130 ACATCAAGTTTCATCAGCAGACTCTCGCGCGCATGCGAGGAGGAGGAGGCTGGCTGAGC 4189  
 424 ysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerA 444  
 4190 AGCAGCGGCGAGGAGCGCGAGCGCTGGCCCGAGGTGGAGGTGGAGAGCAGC 4249  
 444 laArgSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaG 464  
 4250 GGCAGCTGGCGGAGGCGCAGCCCGAGGCAAGCAGCAGCGGAGCGG---GAGCGAAGG 4306  
 464 lnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrSerIleG 484  
 4307 AGCTGACAGCAGCGCATGTCAGAGAGGTGTGGCGCGGAGGAGGCGCGGTGCGAGCGCG 4366  
 484 lnProSerValLysLeuTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProA 503  
 4367 AGCAGCAGAAGCGCAGCATTCAGAGGAGGTGCGCAGCAGCTGCGGCGAGCTCGGAGCGG 4426

503 spProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspG 523  
 4427 AGATCCAGGCCAAGGCCCGGAGCGAGCGGGGTGAGCGCGGCTGCGCATCTCAGG 4486  
 523 lu-----GlnAlaValAspArgArgHisGlnAlaAspPheL 534  
 4487 AGGAGATCCCGCTGGTGGCTCGCTCAGTTGGAGGCCACCGAGCGCCAGGTGGCGGGCTG 4546  
 534 ysLysTrpLeuAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn- 553  
 4547 AGGGGAGCTGCGAGGCACCTCGTGCAGCGCGAGGAGGTGAGGCACAAAAGCGACAGG 4606  
 554 --LysIleAspValGlyLysLeuPheAsnGluVal----- 564  
 4607 CGCAGGAGGAGCGGAGCGCTTGGAGGCGAGGTGCGAGGCGAGCGACGCTAAGCGCGC 4666  
 565 ----ArgAsnLysGluLeuValValAlaProThrLysGluG 577  
 4667 AGCGGAGGTGGAGCTGGCTCGCGCTGAAGCCGAGACCGAGCGCGCGCGCGAGAGC 4726  
 577 luGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThr-----L 591  
 4727 AGCGGCGCTGCGAGGCGCTGCGGCTGCGAGGCGAGGCGGAGCGCGCGCGCGCC 4786  
 591 euArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys----- 608  
 4787 TCGCGCAGGCGGAGGTGGAGCGCGCGGCGAGTACAGGTGGCGCTGGAGCGCGCGCAGC 4846  
 609 ----SerLysValAlaValTyrValAsnLysGlnAlaLeuA 621  
 4847 GCAGTCGAGAGCGGAGCTGCAGAGCAACGCGCTCTTCGCGCGAGACGCGCACAGC 4906  
 621 rgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluL 641  
 4907 TGGAGCGCTCCCTGCGAGGAGAAACAGCTGGCTGTGGCACAGCTGCGGAGAGGCTGAG- 4965  
 641 euLeuLeuCysPheAsnProLeuTriLeuArgLeuGlyLeuGluValValPheGlyGluL 661  
 4966 ----CGCGGGCGACAGCAGCAGCGCGCGCGCGCGCGCGCAGC 4996  
 661 ysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu-----A 679  
 4997 GGGCGCGCGAGGAGCGAGCGGAG-----CTGGAGCGCTGGCAGCTCAAGGCCA 5047  
 679 snArgLeuPheArgAsnLysCysGluGlnArgTyrSerLysAlaTyrThrLeuThrG 699  
 5048 ACAGGCGCTACGCTGCGCTGCGCGGAGGAGGTGGCGCAGCAGCAGAGAGCTGCGCGC 5107  
 699 luGluTyrAlaGluThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProp 719  
 5108 AGCCCGAGGCTGAG----- 5121  
 719 heLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysL 739  
 5122 ----AAGCAGAGGAG----- 5133  
 739 ysSerProHisLysGluThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaA 759  
 5133 ---- 5133  
 759 snIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHis-----A 777  
 5134 ----GAGCGGAGCTCC 5185  
 777 rgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgA 797  
 5186 GGCAGCGGAGCTGGCTGACAGAGCTGAGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 5245  
 797 spGlyValArgLeuThrArgValValGluValIleLeuLeuArg-----A 812  
 5246 CGCAGCAGCGCTGCGCGCGGAGCAGGAGTTCATCCGCTGCGCGCGCGCGCGCGCGCGCG 5305  
 812 spAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnV 832

Db 5306 GGGAGCAGCGCGAGCTGTGGAGGAGGAGCTGGCCCGCTGCAGCGT----- 5355  
 Qy 832 allYsLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPhelGlnLeuGlyClyAspIleAlaA 852  
 Db 5356 -----GAGCGCGCTGCAGCCAGCAGAAACGCGAGAGCTGGAAGCCGAGCTGGCC- 5406  
 Qy 852 laGlnAspIleValAspGlyHisArgGluYsThrLeuSerLeuLeuTrpGlnLeuIleT 872  
 Db 5406 ----- 5406  
 Qy 872 yrLYsPheArgSerProLYsPheHisAlaAlaThrValLeuGlnLYsTrpTrpArgA 892  
 Db 5407 --AAGTGGCGCGCGAG----- 5421  
 Qy 892 rgHisTrpLeuHisValValIleGlnArgArgIleArgHisLYsGluLeuMetArgArgH 912  
 Db 5422 -----ATGGAGGTGTGTGTGCCAAGCAGCGGAGGCTGAGGAGAGTGGCGCTCC- 5472  
 Qy 912 isArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLYsTrpValL 932  
 Db 5472 ----- 5472  
 Qy 932 ysLeuPheLYsThrGluArgThrGlnAlaAlaIleIleLeuGlnLYsPheThrArgArgT 952  
 Db 5473 -----ACCAGCGAGAAGTCCAAGCAGAGAGCTGGAGCGCGAGCGCGCTTCGCGC 5524  
 Qy 952 yrLeuAlaGlnLYsGlnLeuTrpGlnSerLYsHisSerIleIleThrIleGlnArgTrpT 972  
 Db 5525 AGCTGGCGCGAGGAGCGCC----- 5544  
 Qy 972 rpArgAlaGlnLeuGlyArgGln---HisArgGlnArgPheValGluLeuArgGluA 991  
 Db 5545 ---CGCTCGTGCCTTGGCGAAGAGGCGCAAGCGCGAGCGG-----CAGCTGGCGGAG 5596  
 Qy 991 laAlaIlePheLeuGlnArgIleTrpArgArgLeuPheAlaLYsLYsLeuLeuAla 1011  
 Db 5597 AAGACGCGCGCGCGAGCGCGAGCGCGAGCGGGTCTTGGCGAGAAGCTGGCGGCCA 5656  
 Qy 1011 laAlaGluThrAlaArgLeuGlnArgSerGlnLYsGlnAlaAlaAlaSerLYsIleG 1031  
 Db 5657 TCGCGAGCGCACCGCGCTC-----AAGACGAGCGCGAGATCGCGCTCAAGG 5704  
 Qy 1031 lnMetGlnTrpArgThrLYsGlnLeuGlyArgIleGlnArgHisGluPheLeuArgGlnA 1051  
 Db 5705 AGAAGGAGCGGAGAACGAGCGCTTGGCGGCTGGCGGAGACGAGCGCTTCCAGCGCG 5764  
 Qy 1051 rgAspLeuIleMetPheValGlnArgArgMetArgSerLYsTrpSerMetLeuGluGlnA 1071  
 Db 5765 GCGGCTG-----GAGGAGCAGCGCGCGCAACACAAAGCTGCATCGAGGAGC 5812  
 Qy 1071 rgLYsGluPheGlnGlnLeuLYsArgAlaAlaIleAsnIleGlnGlnArgTrpArgAlaL 1091  
 Db 5813 GC-----CTGCGCCAGCTGCGCAAGGATCGAGCAGCGAGCTGGAGCGCGAGAGGGGC 5866  
 Qy 1091 ysLeu-----SerMetArgLYsCysAsnAlaAspLYsLeuAlaLeuArgS 1106  
 Db 5867 TGGTGGAGCACCGCTGAGGAGCGCGCGAGTGGAGAAAGATCTCTGGCGCTGAAGG 5926  
 Qy 1106 erSerValLeuLYsValGlnAlaTrpArgLYsAlaThrIleGlnMetArgIleAspArgA 1126  
 Db 5927 CGAGCTTCGAGAAGCGCGCGCT---GGCAAGCGCGAGCTGAGCTGGAGCTGGGACGC- 5982  
 Qy 1126 snHisLYsTrpSerLeuArgLYsAsnVal---IleCysLeuGlnGlnArgLeuArgAlaI 1145  
 Db 5983 -----ATCCGCACACGCGGAGACACGCTGCGAGCAGAGGAGGAGCGCGC 6028  
 Qy 1145 leMetLYsMetArgGluGlnArg-----GluAsnTrpLYsLeuArgLeuArgAsnA 1161  
 Db 6029 AGCTGGAGCGCGAGGAGCGGAGCTGGCGGCGGAGGAGGAGCGCGCGCGCTGAGG 6088  
 Qy 1161 laSerIleLeuValGlnLYsArgTrpArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181

Db 6089 CTGAGGAGCGCGTGCAGAAAGACCTTGGCGCGCGAGGAGGAGCGCCGACGCGGAGAGG 6148  
 Qy 1181 laTrpLeu-----ArgThrArgLYsCysIleIleAsnValGlnArgArgTrpArgA 1198  
 Db 6149 CGGCGCTGGAGAACTCGAGCGGCTGAAGCAACGCTGGAGGAGGAGCGCGCGCTGCGGG 6208  
 Qy 1198 laThrLeuGlnMetArgArgGluArgLYsAsnTrpLYsHisLeuGlnTrpThrLYsA 1218  
 Db 6209 AGCGAGCGGAGCAGGAGTGGCGCGCGAGCTGCAGTGGCCGAGGAGCGCCGCGCAAGAC 6268  
 Qy 1218 rgIleGlnIleLYsPheArgAlaLYsArgGluMetLYsLYsGlnArgAlaGluPheLeuG 1238  
 Db 6269 GGCTGCAGCGCGAGAGAAAGCACAACGCTTGGCGGTGCAGCAGAAAGAGAGAGAGCTTAC 6328  
 Qy 1238 lnLeuLYsLYsValThrLeuValGlnLYsArgArgAlaLeuLeuGlnMetArgL 1258  
 Db 6329 AG-----CAGACGCTGCAGAGGAGCAGAGCGTGTGGACGAGCTGCGCG 6373  
 Qy 1258 ysGluArgGlnGluTrpIleuHisLeuArgGluValThrIleLYsLeuGlnArgPheH 1278  
 Db 6374 CGGAGCGCGAG-----CGCGCGCGCGCGCGCGG 6400  
 Qy 1278 isAlaGlnLYsSerMetArgPheMetArgAlaLYsTrpArgGlyThrGlnAlaAlaValS 1298  
 Db 6401 CTGAGGAGCGGAGGAGCGCGGTGCGAGCGGAGCTGAGCGCGCGAGGCG----- 6453  
 Qy 1298 erCysLeuGlnMetHisTrpArgAsnHisLeuLeuArgLYsArgGluArgAsnSerPheL 1318  
 Db 6454 -----CGCGCGCAGGTGGAAGGCGCG 6475  
 Qy 1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTrpArgAlaArgLeuAsnMetI 1338  
 Db 6476 AGCGCTGAAGCAGCTCGGCA-----G 6496  
 Qy 1338 leLYsGlnLeuLYsSerTrpAlaGlnLeuLYsGlnAlaAlaIleThrIleGln----- 1355  
 Db 6497 AGGAGCAGGACAGCGCGCGCTCAGGCACAGCGCGCTGCAGAGAGTCTGGCAAGAGGAG 6556  
 Qy 1356 --ThrArgTrpArgAlaLYsLYsAlaMetGlnLYsGlnValValLeuTrpGlnLYsGln- 1374  
 Db 6557 CGGAGCAGAGCGCGCGCGCGCGCACAGCGGAGCAGCGCGCTTGGCGGAGAGCAGG 6616  
 Qy 1375 --ArgGluAlaIleIleLYsValGlnArgArgTrp---ArgGlyAsnLeuGluMetArgL 1393  
 Db 6617 CAGCTGCAGCGAGATGGAGAAAGATAAGAAATTCGCGAGCAGCAGCTGGCGGCAAGAG 6676  
 Qy 1393 ysGlnIleGluValTrpGlnLYsGlnArgGlnAlaValIleArgLeuGlnLYsTrpTPA 1413  
 Db 6677 CGCAGGTGGAG-----CAGGAGCTGACACACTCGCGGTGCAG----- 6714  
 Qy 1413 rgSerIleArgAspMetArgLeuCysLYsAlaGlyTrpArgArgIleArgLeuSerSerL 1433  
 Db 6715 -----CTGGAGGAGACCGACACAGAG---AACCTGTCTGGAGC 6751  
 Qy 1433 euSerIleGlnArgLYsTrpArgAlaThrValGlnAlaArgArgGlnArg----- 1449  
 Db 6752 AGGAGCTGCAGCGCTGAAGCGGAGGCGCACCGAGGCGCGCACCGCAGCGAGCCAGCTGG 6811  
 Qy 1450 ----GluIlePheLeuSerThrIleArgLYsValArgLeu-----MetGlnAlaPheI 1466  
 Db 6812 AGGAGAGAGCTTCTTCGTGGCGGTGCAGATGGAGGAGCTGAGCAAGCTCAAGGACCGCA 6871  
 Qy 1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnArgArgGluP 1479  
 Db 6872 TCGAGCTGAGAACCGCGCACTCATCTTGGCTGACAGGACAAATACGAGCGGCTTCTTGC 6931  
 Qy 1479 heGluMetLYsArgArgAlaAlaValValIleGlnArgArgPheArgAlaArgCysAlaM 1499  
 Db 6932 AGGAGAGGCTGAGAAGATGAAGCAGGTGGCGGAGGAGCGCGCGGTGAGCTGTGGCGG 6991  
 Qy 1499 etLeuLYsAlaArgGlnAspTrpGlnLeuIleGlnSerValIleLeuGlnArgL 1519  
 Db 6992 CCCAAGAGGCTCGCGCAGCTGCGGAGTGGCAGAGGAGAC-----CTGGCACAGCAGC 7045



3833 AGGCCATGCGCTGGCCGACACAGCCAGGCTGTGCGGAGCAGCTGCGGAGGAGGCGCC 3892  
Qy  
360 -----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe----- 375  
Db  
3893 TGTGGAGGAGATCGAGCGCCAGCGGAGAGAGGTCTGAGAGGTGCGAGAGGTTTGGAAAC 3952  
Qy  
376 -----AsnLeuHisGluValGlyArgLysSerValLysG 387  
Db  
3953 AGTACATCAACGCCATCAAGGACTATGAATCCAGCTCGTGTGACGTACAGGCG--CAGC 4009  
Qy  
387 lySerProValLysAsnProHisLysArg----- 396  
Db  
4010 TTGAGCGGTGCGCTCCCGGCCAAGAAAGCCAGGTCAGTCCAGGATCATGAGAGTGTCA 4069  
Qy  
397 -----ArgSerHisGluLeuSerPheSerAspAlaProSer---- 408  
Db  
4070 TCCAGGAGTACGTGGAACCTGCGTACGCACCTACAGGAGCTACCACTGAGGAGCCAGT 4129  
Qy  
409 -----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProL 424  
Db  
4130 ACATCAAGTTTCATCAGCGAGACTCTGCGCGCATGAGGAGGAGGCTGGCTGAGC 4189  
Qy  
424 yLysGlnArgValGlnAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerA 444  
Db  
4190 AGCAGCGGCAGAGGAGCGCGAGCGGCTGGCGAGGTGGAGCGCGCTGGAGAACGAGC 4249  
Qy  
444 laArgSerSerSerAlaHisAlaTTPProHisAlaGlnSerLysLysPheLysLeuAlaG 464  
Db  
4250 GGCGAGCTGGCGAGCGGACGCCAGGCAAGGCAAGGCGAGGAGCGG--GAGGCGAAGG 4306  
Qy  
464 lnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrSerIleG 484  
Db  
4307 AGCTGCACGCGCATGACGAGGAGGAGTGTGCGCGGAGGAGGCGCGGTGGAGCGCG 4366  
Qy  
484 lnProSerValLysLeuTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProA 503  
Db  
4367 AGCAGAGAGCGCAGCATTCAGGAGGAGCTGCAGCAGCTGCGGAGAGCTGGGAGCGG 4426  
Qy  
503 spProPheAlaAlaThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspG 523  
Db  
4427 AGATCCAGGCCAAGGCCCGGAGGAGGAGGCGGCTGAGGCGAGCGCGCTGCGCATCGAGG 4486  
Qy  
523 lu-----GlnAlaValAspArgHisGlnAlaAspPheL 534  
Db  
4487 AGGAGATCGCGTGTGCGCTGCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4546  
Qy  
534 yLysTyrPheAsnAlaLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn- 553  
Db  
4547 AGGGGAGCTGCAGGCACTGCGTGCACGGCGGAGGAGGCTGAGGCAAAAGACGACAGG 4606  
Qy  
554 --LysIleAspValGlyLysLeuPheAsnGluVal----- 564  
Db  
4607 GCAGAGGAGGCCGAGCGCTTGGGAGGCGAGGTGCAGGACGAGAGCCAGCGTAAGCGGC 4666  
Qy  
565 -----ArgAsnLysGluLeuValValAlaProThrLysGluG 577  
Db  
4667 AGGCGGAGGTGGAGCTGCGCTGCGCGTGAAGCGCGAGACCGAGCGCGCGCGAGAGC 4726  
Qy  
577 luGlnSerMetAsnTyrLeuThrLysTyrArgLeuGluThr-----L 591  
Db  
4727 AGCGGGCCCTGCAGGCCCTGGAGGAGCTGCGGCTCAGCGCGAGGAGGCGGAGCGCGCC 4786  
Qy  
591 euArgLysAlaAlaValGluLeuPhePheSerGluGlnMetArgLeuProCys----- 608  
Db  
4787 TGGCGAGGCCGAGGTGGAGCGAGCGCGGAGGTACAGTGGCCCTGGAGAGCGGCGAGC 4846  
Qy  
609 -----SerLysValAlaValTyrValAsnLysGlnAlaLeuA 621  
Db  
4847 GCAGTGCAGAGCGGAGCTGCAGAGCAAGCGGCTCTTCGCCGAGAGAGCGGACGACAGC 4906  
Qy  
621 rgIleArgSerAspArgAsnLeuHisLeuAspValValMetGlnArgThrIleLeuGluL 641  
Db  
4907 TGGAGCGCTCCCTGCGAGGAGGAGACAGTGGCTGTGGCAGCTGCGGAGGAGGCTGAG- 4965

641 euLeuLeuCysPheAsnProLeuTyrPheArgLeuGlyLeuGluValValPheGlyGluL 661  
4966 -----CGGCGGCACAGCAGCAGCGCGCGGAGC 4996  
Qy  
661 ysIleGlnMetGlnSerAsnArgAspIleValGlyLeuSerThrPheIleLeu-----A 679  
Db  
4997 GGGCGCGAGGAGGAGCGAGCGGAG-----CTGGAGCGCTGGCAGCTCAAGGCCA 5047  
Qy  
679 snArgLeuPheArgAsnLysCysGluGluGlnArgTyrSerLysAlaTyrThrLeuThrG 699  
Db  
5048 ACAGAGCGCTACGCGTGGCTGCGCGGAGGAGGTGCGCAGCAGAGAGCTTGGCGC 5107  
Qy  
699 luGluTyrAlaGlnThrIleLysLysHisSerLeuGlnLysIleLeuPheLeuLeuProP 719  
Db  
5108 AGGCGGAGCTGAG----- 5121  
Qy  
719 heLeuAspGlnAlaLysGlnLysArgIleValLysHisAsnProCysLeuPheValLysL 739  
Db  
5122 -----AAGCAGAAGGAG----- 5133  
Qy  
739 ysSerProHisLysLysThrLysAspIleLeuLeuArgPheSerSerGluLeuLeuAlaA 759  
Db  
5133 ----- 5133  
Qy  
759 snIleGlyAspIleThrArgGluLeuArgArgLeuGlyTyrValLeuGlnHis-----A 777  
Db  
5134 -----GAGCGGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 5185  
Qy  
777 rgGlnThrPheLeuAspGluPheAspTyrAlaPheAsnAsnLeuAlaValAspLeuArgA 797  
Db  
5186 GCGCAGCGGAGCTGGCTGAACAAGAGCTGGAGAGCAGCAGCGGAGCTGCGGAGGAGCAGC 5245  
Qy  
797 spGlyValArgLeuThrArgValValGluValIleLeuLeuArg-----A 812  
Db  
5246 CCGCAGCAGCGCTGCGCGCGAGCGAGGAGTGTATCCGCTGCGGCGCGAGCAGGAGCAGG 5305  
Qy  
812 spAspLeuThrArgGlnLeuArgValProAlaIleSerArgLeuGlnArgIlePheAsnV 832  
Db  
5306 GGGAGCAGCAGCGCGAGCTGTGGAGGAGGAGCTGCGCGCGTGCAGCGT----- 5355  
Qy  
832 allYsLeuAlaLeuGlyAlaLeuGlyGluAlaAsnPheGlnLeuGlyGlyAspIleAlaA 852  
Db  
5356 -----GAGCGCGCTGCAGCCAGCAGAAACCGCAGAGAGCTGGAACCGCGAGTGGCC- 5406  
Qy  
852 laGlnAspIleValAspGlyHisArgGluLysThrLeuSerLeuLeuTyrGlnLeuIleT 872  
Db  
5406 ----- 5406  
Qy  
872 yrLysPheArgSerProLysPheHisAlaAlaAlaThrValLeuGlnLysTyrTrpArgA 892  
Db  
5407 --AAGGTGCGGCGCGAG----- 5421  
Qy  
892 rgHisTrpLeuHisValValIleGlnArgArgIleArgHisLysGluLeuMetArgArgH 912  
Db  
5422 -----ATGAGAGTGTCTGCTGCACCAAGCGGAGGCTGAGAGAGTGCCTGCTCC- 5472  
Qy  
912 isArgAlaAlaThrValIleGlnAlaValPheArgGlyHisGlnMetArgLysTyrValL 932  
Db  
5472 ----- 5472  
Qy  
932 ysLeuPheLysThrGluArgThrGlnAlaAlaIleLeuGlnLysPheThrArgArgT 952  
Db  
5473 -----ACCAGCGAGAAGTCCAAGCAGAGAGCTGGAGCGCGCGCGCGTTCGCGG 5524  
Qy  
952 yrLeuAlaGlnLysGlnLeuTyrGlnSerTyrHisSerIleIleThrIleGlnArgTyrT 972  
Db  
5525 AGCTGCGCGAGGAGCGCGC----- 5544  
Qy  
972 trArgAlaGlnGlnLeuGlyArgGln---HisArgGlnArgPheValGluLeuArgGluA 991  
Db  
5545 --CGCTGCGTCCCTGCGGAGGAGGAGCGCGAGCGCGCGG-----CAGCTGGCGGAGG 5596

QY 991 laAlaIlePheLeuGlnAArgIleTrrArgArgLeuPheAlaLysLysLeuAlaA 1011  
 Db |||||  
 QY 5997 AAGACGGCGCGCAGCGCGCCAGCGCGGCTTTCGCGAAGAGCTGCGCGCA 5656  
 Db |||||  
 QY 1011 laAlaGluThrAlaArgLeuGlnAArgSerGlnLysGlnAlaAlaSerTyrIleG 1031  
 Db |||||  
 QY 5657 TCGCGAGCGCCACGCGCTC-----AAGACGGAGCGGAGATCGCGCTCAAGG 5704  
 Db |||||  
 QY 1031 lnMetGlnTrrArgTrrTyrGlnLeuGlyArgIleGlnAArgHisGluPheLeuArgGlnA 1051  
 Db |||||  
 QY 5705 AGAAGGAGCGGAGAAACAGCGCTCGCGCGCTGCGCGAGAGAGAGAGCTTCACGCGG 5764  
 Db |||||  
 QY 1051 rgAspLeuIleMetPheValGlnAArgMetArgSerLysTrrSerMetLeuGlnGlnA 1071  
 Db |||||  
 QY 5765 GCGCGCTG-----GAGGAGCAGCGCGGCAACAGAGCTGCATCGAGGAGC 5812  
 Db |||||  
 QY 1071 rgLysGluPheGlnGlnLeuLysArgAlaIleAsnIleGlnGlnAArgTrrArgAla 1091  
 Db |||||  
 QY 5813 GC-----CTGGCCAGCTGCGCAAGCATCGGACAGCAGCTGGAGCGGCAAGGGCG 5866  
 Db |||||  
 QY 1091 ysLeu-----SerMetArgLysCysAsnAlaAAspTrrLeuAlaLeuArgS 1106  
 Db |||||  
 QY 5867 TGTGTGAGGACACCTGAGGCGCGCGGAGGTGGAGAGAGATCTCGCGCTGAAGG 5926  
 Db |||||  
 QY 1106 erSerValLeuLysValGlnAlaTrrArgLysAlaThrIleGlnMetArgIleAspArgA 1126  
 Db |||||  
 QY 5927 CGAGCTTCGAGAGCGCGCT---GGCAAGGCGAGCTGGAGCTGGAGTGGAGCGC- 5982  
 Db |||||  
 QY 1126 snHisTrrTyrSerLeuArgLysAsnVal---IleCysLeuGlnGlnAArgLeuArgAla 1145  
 Db |||||  
 QY 5983 -----ATCCGACGAAACGCGGAGGACACGCTGCGCAGCAAGGAGCAGCGCG 6028  
 Db |||||  
 QY 1145 leMetLysMetArgGlnGlnAArg-----GluAsnTrrLeuArgLeuArgAsnA 1161  
 Db |||||  
 QY 6029 AGCTGTGAGCGCGCAGGACGCGCAGCTGCGCGGAGGAGGAGCGCGCGCTGAGG 6088  
 Db |||||  
 QY 1161 laSerIleLeuValGlnLysArgTrrArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181  
 Db |||||  
 QY 6089 CTGAGGAGCGCTGCAGAGAGCTGCGCGCGGAGGAGGCGCGCAGCGGAGG 6148  
 Db |||||  
 QY 1181 laTrrLeu-----ArgThrArgLysCysIleIleAsnValGlnAArgTrrArgA 1198  
 Db |||||  
 QY 6149 CGCGCTGTGAGGAAGTGCAGCGGCTGAAAGCAACGTGGAGGAGCGCGCGCTGCGG 6208  
 Db |||||  
 QY 1198 laThrLeuGlnMetArgArgGluArgLysAsnTrrLeuHisLeuGlnThrThrLysA 1218  
 Db |||||  
 QY 6209 AGCGAGCGGAGGAGTTCGCGCGGCGCAGCTGCGCTGGCCCGCAGGAGCGCGCCAGAG 6268  
 Db |||||  
 QY 1218 rgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnAArgAlaGluPheLeuG 1238  
 Db |||||  
 QY 6269 GGCTGCAGCGGAGAGAGAGCAGCGCTTCGCGGTGCAGCAGAGAGGAGGAGGCTAC 6328  
 Db |||||  
 QY 1238 lnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgL 1258  
 Db |||||  
 QY 6329 AG-----CAGACGCTGCAGCAGGAGCAGCGCTGCTGGACCACTGCTGCGG 6373  
 Db |||||  
 QY 1258 ysGluArgGlnGluTrrLeuHisLeuArgGluValThrIleLysLeuGlnAArgPheH 1278  
 Db |||||  
 QY 6374 GCGAGCGGAG-----GCGGCGCGCGCGCGG 6400  
 Db |||||  
 QY 1278 isAlaGlnLysSerMetArgPheMetArgAlaLysTrrArgGlyThrGlnAlaAlaVal 1298  
 Db |||||  
 QY 6401 CTGAGGAGCGGAGGAGCGCGGTGCAGCGGAGCGCTGAGCGCGCGCGCGGCGC- 6453  
 Db |||||  
 QY 1298 erCysLeuGlnMetHisTrrArgAsnHisLeuLeuArgLysArgGluArgAsnSerPheL 1318  
 Db |||||  
 QY 6454 -----CGCGCGCAGGTGGAAGAGCGCG 6475  
 Db |||||  
 QY 1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnAArgTrrArgAlaLeuAsnMetI 1338  
 Db |||||  
 QY 6476 AGCGCTGAGAGCTGCGCA-----G 6496  
 Db |||||  
 QY 1338 leLysGlnLeuLysSerTrrAlaGlnLeuLysGlnAlaAlaIleThrIleGln----- 1355  
 Db |||||

Db 6497 AGGAGCAGGCACAGCGCGGCTCAGCAGCAGCGCGCTGCAGAGAGAGTGCAGAGGAGG 6556  
 QY 1356 --ThrArgTrrArgAlaLysLysAlaMetGlnLysGlnValValLeuTrrGlnLysGln- 1374  
 Db |||||  
 QY 6557 CCGAGCAAGAGCGCGCGCGGCGCAGCGGAGCAGCGCGCTTGCAGAGAGCAGG 6616  
 Db |||||  
 QY 1375 --ArgGluAlaIleIleLysValGlnAArgTrr---ArgGlyAsnLeuGlnMetArgL 1393  
 Db |||||  
 QY 6617 CAGCTGACGCGGAGATGAGAAATTCGCGCAGCAGCAGCTGCGCGGAGAGG 6676  
 Db |||||  
 QY 1393 ysGlnIleGluValTrrGlnLysGlnAlaValIleArgLeuGlnLysTrrTrpA 1413  
 Db |||||  
 QY 6677 GCGAGGTGAG-----CAGGAGCTGACACACTGCGGCTGAG----- 6714  
 Db |||||  
 QY 1413 rgSerIleArgAspMetArgLysAlaGlyTrrArgArgIleArgLeuSerSerL 1433  
 Db |||||  
 QY 6715 -----CTGAGGAGACCGACACACAGAG---AACCTGCTGGAGC 6751  
 Db |||||  
 QY 1433 euSerIleGlnAArgLysTrrArgAlaThrValGlnAlaAArgArgGlnAArg----- 1449  
 Db |||||  
 QY 6752 AGGAGCTGCAGCGGCTGAAGCGGAGCGCCAGCGGCGCGCAGCAGCAGCGCAGGTGG 6811  
 Db |||||  
 QY 1450 ----GluIlePheLeuSerThrIleArgLysValArgLeu-----MetGlnAlaPheI 1466  
 Db |||||  
 QY 6812 AGGAGGAGCTCTTCTCGTGCCTGCGATGAGGAGCTGAGCAAGCTCAAGGACGCA 6871  
 Db |||||  
 QY 1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnAArgArgGluP 1479  
 Db |||||  
 QY 6872 TCGAGGCTGAGAACCGCGCACTCATCTGCTGACAGGACATACCGAGGCTTCCTGC 6931  
 Db |||||  
 QY 1479 heGluMetLysArgArgAlaValIleGlnAArgPheArgAlaAArgCysAlaM 1499  
 Db |||||  
 QY 6932 AGGAGGAGCTGAGAGATGAAGCAGGTGCGGAGGAGCGCGCGCTGAGTGTGCGG 6991  
 Db |||||  
 QY 1499 etLeuLysAlaArgGlnAspTrrGlnLeuIleGlnSerSerValIleLeuValGlnA 1519  
 Db |||||  
 QY 6992 CCCAAGGCTGCGCACTGCGGAGCTGCGCAGCAGAGGAGGAGC-----CTGGCAGCAGC 7045  
 Db |||||  
 QY 1519 ysPheArgAlaAsnArgSerMetLys-----GlnAlaArgGlnGluPheValGlnL 1536  
 Db |||||  
 QY 7046 GGGCTTGGCAGAGAGATGCTCAGGAGAGAGATGCGGCGGTGCAGAGGCGCACGCGAC 7105  
 Db |||||  
 QY 1536 euArgThrIleAlaValHisLeuGlnGln-----LysPheArgGlyLysA 1551  
 Db |||||  
 QY 7106 TCAAGCTGAGCGGAACTGCTCAGCAGCAGAGAGGAGCTTGGCGCAGGAGCGCGCG 7165  
 Db |||||  
 QY 1551 rgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheG 1571  
 Db |||||  
 QY 7166 GGCTGCAGGAGCAGAGCAGATGCGCGCAGCTTGGCGGAGGAGACCGCAGGCTTCC 7225  
 Db |||||  
 QY 1571 lnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGlu- 1590  
 Db |||||  
 QY 7226 AG-----CGACGCTGAGCGCGCAGCGCGCAGCTGGAGATGAGCGCTGAGG 7276  
 Db |||||  
 QY 1591 -----MetMetAspLeuIleArg----- 1596  
 Db |||||  
 QY 7277 CTGAGCGCTCAAGCTGCTGCGCGAGATGAGCGCGCAGCGCGCGCTGAGGAGG 7336  
 Db |||||  
 QY 1597 -----GlnLysArgAlaAlaLysValIleGlnAArgTrrTrpArgGlyTrrL 1612  
 Db |||||  
 QY 7337 ACGCCAGCGCTTCCGGAAGAGCGGAGGAGATCGGTGAGAGCTGCACCGCAGCGAGC 7396  
 Db |||||  
 QY 1612 eu-----IleArgArgArgGlnLysH 1619  
 Db |||||  
 QY 7397 TCGCCACCCAGGAGAGGTGACCTGCTGCAGACACTGGAGATCCAGCGCAGCAGAGTG 7456  
 Db |||||  
 QY 1619 isGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlav 1639  
 Db |||||  
 QY 7457 ACCATGATCCGAGCGCTGCGGAGGCGCATCGTGTGAGCTGAGCGCTGAGAGAGG--- 7512  
 Db |||||  
 QY 1639 alAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleA 1659  
 Db |||||

Db 7513 -----AAGCTCCACAGCAGGCGCAAACTGCTGCAG-----CTCA 7546  
Qy 1659 laSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrVal-ProHis 1678  
Db 7547 AGTCTGAGGATGACAGCGTGCAGCAGGAGCAGCTGCTCAGGAGACGAGGCC--- 7603  
Qy 1679 LeuLeuMetTrpCysSerGluPheMetSer-----ThrPheCysTyrGlyMet 1695  
Db 7604 -----TGCAGCAAGACTTCCTCTCTGAAAGGACAGCTGCTACAGCGGAGC 7651  
Qy 1696 AlaGlnAlaLeuArg 1700  
Db 7652 GCTTCATCGACGAG 7666

## RESULT 14

US-10-717-597-183  
; Sequence 183, Application US/10717597  
; Publication No. US20040110221A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael E.  
; APPLICANT: Twine, Natalie C.  
; APPLICANT: Dorner, Andrew J.  
; APPLICANT: Trepicchio, William L.  
; APPLICANT: Slonim, Donna K.  
; APPLICANT: Stover, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS  
; FILE REFERENCE: AM101080L  
; CURRENT APPLICATION NUMBER: US/10/717,597  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: US 60/459,782  
; PRIOR FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: US 60/427,982  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 4904  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 183  
; LENGTH: 14800  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-717-597-183

Alignment Scores:  
Pred. No.: 7,8e-20 Length: 14800  
Score: 322.00 Matches: 363  
Percent Similarity: 35.69% Conservative: 289  
Best local Similarity: 19.87% Mismatches: 682  
Query Match: 3.38% Indels: 493  
DB: 17 Gaps: 72

US-09-914-698-1 (1-1861) x US-10-717-597-183 (1-14800)

Qy 77 LysLysArgMetSerAlaAlaAlaProProSerLysGlnThr-----Trp 92  
Db 3062 AGGAAGAGTCTCGTGCAGCGCTCATCTCCGAGCTCAAAGACATCCGGCTGCAGCTGG 3121  
Qy 93 ArgValThrAlaProSerArgProAlaAlaTrpAla-HisProProGlnAla-ProL 112  
Db 3122 -----AGGCTGTGACACGCGCACCGTGCACCGCTCGCGCTGCCGC 3163  
Qy 112 euValGluLysAsnValTyrLysThrProGlnGluProValTyrLysSerProGlnP 132  
Db 3164 TGGACAAAGCGCGCAGCGGAGTGTGCCAGCGC-----ATCGCCGAGCAGC 3211  
Qy 132 roArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuAspValIleA 152  
Db 3212 AGAAGGCACAGGCAGAG----- 3228  
Qy 152 spAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyGlnAlaThrIlePhe- 171  
Db 3229 -----GTGAGGGGCTGGCGAAGGGGTGCGCCGCGCTCTCG 3265  
Qy 172 -----ProAspAsnLeuAlaAlaTrpPro-----ThrProThrLeuLysG 185

Db 3266 CCGAGCGCGAGAGGTCTTGGCCCTACAGAGCCATCGCTGCGGCCCCCGCTGGCGCT 3325  
Qy 185 lyAsnValLysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuG 205  
Db 3326 CGGAGCTGGAG-----CTGACGCTGGCGAGCTGG 3355  
Qy 205 luAspGlnProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnIleS 225  
Db 3356 AGCAGGTCGCGAGCTGTCTGCCATCTAC---CTGGAGAAGCTCAAGACCATCAGCTGG 3412  
Qy 225 erLeuAspThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysT 245  
Db 3413 TGATCCGCGGCACGACGAGCGCGGAGGAGTGTCTCAGGCGCCACAGAGCAGCTCAAG 3472  
Qy 245 hrThrThrIleValHisAlaThrHisThrArg-----AlaLeuAlaCysIleHisG 262  
Db 3473 AGGCCCGAGCGCTGCCGCGCCACCTCCCGAGCTCGAGGCCACCAAGCCCTCTGAAGA 3532  
Qy 262 luGluGluGlyProSerProProArgThrProThrLysSerAlaIleHisAspLeuLysA 282  
Db 3533 AGCTGGGGCCCGAGCGCGGAGGCACAGCAGCCACGTTCCAGCGCCCTGGCGGATGAGCTGC 3592  
Qy 282 rgAspIleLysLeuValGlySerProLeu---ArgLysTyrSerGluSerMetLysAspL 301  
Db 3592 GGGGGGCACAGAGGTGGGGGAGGAGCTGCAGCAGCGGCACGGGAGCGGAGCTTGGAGG 3652  
Qy 301 euSerLeuLeuSerProGlnThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnG 321  
Db 3653 TGGAGCGCTGGCGGAGCGGGTCCGCCAGGTGCTTGAGCGCTGGCAGGCTGTGCTGGCCC 3712  
Qy 321 luMetLysIleArgSerIleGlu-----GlnAsnArgTyrTyrGlnG 335  
Db 3713 AGACCGACTTGGCGCAGCGCGAGCTCGAGCAACTGGGCGCGCAGCTGCTTACTACCGCG 3772  
Qy 335 lu-----GlnGlnIleGlnIleL 341  
Db 3773 AGAGTGCAGACCCCTTGGCGCGCTGGTGCAGGAGCCCGCGCGCGAGGAGAGATCC 3832  
Qy 341 ysAlaLysAspLeuAsnSerSerSerSerSerGluAlaSerLeuAlaGlyGlnGln--- 359  
Db 3833 AGGCATCCCTGCTGGCCGACAGCCAGCTGTGGGGAGCAGCTGCGGAGGAGCAGGCC 3892  
Qy 360 -----GluPheLeuPheAsnHisSerGluIleLeuAlaGlnSerSerArgPhe----- 375  
Db 3893 TGCTGGAGGAGATCGAGCGCCACGCGGAGAAGGTGCGAGAGTGCAGAGGTTTGCAGAAC 3952  
Qy 376 -----AsnLeuHisGluValGlyArgLysSerValLysG 387  
Db 3953 AGTACATCAACGCCCATCAAGGACTAAGACTTCACTCCAGCTGGTGACGTACAGGCG---CAGC 4009  
Qy 387 lySerProValLysAsnProHisLysArg----- 396  
Db 4010 TTGACCGCGTGGCTCCCGGCCAAGAGCCCAAGGTCCAGTCGGGATCAGAGAGTGTC 4069  
Qy 397 -----ArgSerHisGluLeuSerPheSerAspAlaProSer----- 408  
Db 4070 TCCAGGAGTACGTGTGACCTGCGTACGCACATACAGCAGCTGACCCACTGACAGCCAGT 4129  
Qy 409 -----AsnGluSerLeuTyrArgAsnGluThrValAlaIleSerProL 424  
Db 4130 ACATCAAGTTTCATCAGCAGAGACTCTCGCGCGCATCGAGAGGAGGAGGCTGGCTGAGC 4189  
Qy 424 ysLysGlnArgValGluAspThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerA 444  
Db 4190 AGCAGCGGCAGAGGAGCGCGAGCGCTGGCCGAGGTGGAGCGCGCTGGAGAGCAGC 4249  
Qy 444 laArgSerSerSerAlaHisAlaTrpProHisAlaGlnSerLysLysPheLysLeuAlaG 464  
Db 4250 GGCAGCTGGCGCGAGCGCAGCCAGGCAAGGCAAGCAGCGAGCGG---GAGCGAAGG 4306  
Qy 464 lnThrMetSerLeuMetLysLysProAlaThrProArgLysValArgAspThrIleG 484





1145 leMetIysMetArgGluGlnArg-----GluAsnTyrLeuArgLeuArgAsnA 1161  
Db : : : : :  
6029 AGCTGAGGCGCGAGCGAGCGAGCTGCGCGGAGGAGGAGCGCGCGCTGAGG 6088  
Qy : : : : :  
1161 laSerIleLeuValGlnLysArgTyrArgMetArgGlnGlnMetIleGlnAspArgAsnA 1181  
Db : : : : :  
6089 CTGAGGAGCGGTGCAGAGAGCGCTGCGCGCGAGGAGGAGCGCGCGAGGAGG 6148  
Db : : : : :  
1181 laTyrIleu-----ArgThrArgLysCysIleIleAsnValGlnArgArgTyrArgA 1198  
Qy : : : : :  
6149 CGCGCGTTCGAGGAGTGCAGCGCTGAAGCGCTGAAGCGCGAGGAGGCGCGCTCGCG 6208  
Db : : : : :  
1198 laThrLeuGlnMetArgArgGluArgLysAsnTyrLeuHisLeuGlnThrThrThrLysA 1218  
Qy : : : : :  
6209 AGCAGCGGAGGAGGAGTGCAGCGCGAGCTGCGCGCGAGGAGGCGCGCGAGGAG 6268  
Db : : : : :  
1218 rgIleGlnIleLysPheArgAlaLysArgGluMetLysLysGlnArgAlaGluPheLeuG 1238  
Qy : : : : :  
6269 GGTGTCAGCGCGAGAGGAGCACCGCTTCGCGGTGCGAGCAGAGGAGGAGGAGCTAC 6328  
Db : : : : :  
1238 lnLeuLysLysValThrLeuValValGlnLysArgArgAlaLeuLeuGlnMetArgL 1258  
Qy : : : : :  
6329 AG-----CAGACGCTGCAGCAGCAGCAGCGCGCTGCGACCGAGCTCGCG 6373  
Db : : : : :  
1258 ysGluArgGlnGluTyrLeuHisLeuArgGluValThrIleLysLeuGlnArgArgPheH 1278  
Qy : : : : :  
6374 GCGAGCGGAG-----GCGCGCGCGCGCGCGG 6400  
Db : : : : :  
1278 isAlaGlnLysSerMetArgPheMetArgAlaLysTyrArgGlyThrGlnAlaValS 1298  
Qy : : : : :  
6401 CTGAGGAGCGGAGGAGCGCGGTGCGCGAGCGGTGAGCGCGCGAGGCG----- 6453  
Db : : : : :  
1298 erCysLeuGlnMetHisTyrArgAsnHisLeuArgLysArgGlnArgAsnSerPheL 1318  
Qy : : : : :  
6454 -----CGCGCGAGGTGGAGGCGCG 6475  
Db : : : : :  
1318 euGlnLeuArgGlnAlaAlaIleThrLeuGlnArgArgTyrArgAlaArgLeuAsnMetI 1338  
Qy : : : : :  
6476 AGCGGTGAAGCAGTGGCA-----G 6496  
Db : : : : :  
1338 leLysGlnLeuLysSerTyrAlaGlnLeuLysGlnAlaAlaIleThrIleGln----- 1355  
Qy : : : : :  
6497 AGGAGGCGACAGGCGCGGCTCAGCGACAGCGGCTCAGAGAGCTGCGCAGAGGAG 6556  
Db : : : : :  
1356 --ThrArgTyrArgAlaLysLysAlaMetGlnLysGlnValValLeuTyrGlnLysGln- 1374  
Qy : : : : :  
6557 CCGAGCAAGAGCGCGCGCGCGACAGCGGAGCAGCGCGCTCGCGCAGAGAGCAGG 6616  
Db : : : : :  
1375 --ArgGluAlaIleLysValGlnArgTyr---ArgGlyAsnLeuGluMetArgL 1393  
Qy : : : : :  
6617 CAGCTCAGCGGAGATGGAGAGCATAGAAATTGCGCGAGCAGCGCTGCGCGCAGAGG 6676  
Db : : : : :  
1393 ysGlnIleGluValTyrGlnLysGlnArgGlnAlaValIleArgLeuGlnLysTyrTrpA 1413  
Qy : : : : :  
6677 CGCAGGTGAG-----CAGAGCTGACACACTGCGGCTGCAG----- 6714  
Db : : : : :  
1413 rgSerIleArgAspMetArgLeuCysLysAlaGlyTyrArgArgIleArgLeuSerSerL 1433  
Qy : : : : :  
6715 -----CTGAGGAGACCGACCCAGAGAG---AACCTGCTGGAGC 6751  
Db : : : : :  
1433 euSerIleGlnArgLysTyrArgAlaThrValGlnAlaArgArgGlnArg----- 1449  
Qy : : : : :  
6752 AGAGCTGACGCGGTGAGGCGGAGCGACCGAGGCGCGACCGCGCAGCGCAGCGAGTGG 6811  
Db : : : : :  
1450 -----GluLePheLeuSerThrIleArgLysValArgLeu-----MetGlnAlaPheI 1466  
Qy : : : : :  
6812 AGGAGGAGCTTCTTCGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6871  
Db : : : : :  
1466 leArgAla-----ThrLeuLeuMetArgGln-----GlnArgArgGluP 1479  
Qy : : : : :  
6872 TCAGGCTGAGAACCGCGGACTCATCTGCGTGACAGGACAAATACGAGCGCTTCCTGCG 6931  
Db : : : : :  
1479 heGluMetLysArgArgAlaAlaValIleGlnArgPheArgAlaArgCysAlaM 1499  
Qy : : : : :

6932 AGCAGGAGCGCTGAGAGATGAAGCAGGTGCGGAGGAGCGCGCGCTGAGTGTGCGG 6991  
Db : : : : :  
1499 etLeuLysAlaArgGlnAspTyrGlnLeuIleGlnSerSerValIleLeuValGlnArgL 1519  
Qy : : : : :  
6992 CCCAAGAGCTTCCGCGAGTCCGCGAGTCCGCGAGTCCGCGAGGAGGAGC-----CTGCGCAGCAGC 7045  
Db : : : : :  
1519 ysPheArgAlaAsnArgSerMetLys-----GlnAlaArgGlnGluPheValGlnL 1536  
Qy : : : : :  
7046 GCGCGCTTGCAGAGAGATGCTCAAGGAGAGAGATGTCAGGCGGTGTCAGAGGAGCGCAGCGAC 7105  
Db : : : : :  
1536 euArgThrIleAlaValIleLeuGlnGln-----LysPheArgGlyLysA 1551  
Qy : : : : :  
7106 TCAGGCTGAGCGGAGAACTGCTGCAGCAGCAGAGAGAGCTTGCAGCAGGAGCGCGCG 7165  
Db : : : : :  
1551 rgLeuMetIleGluGlnArgAsnCysPheGlnLeuLeuArgCysSerMetProGlyPheG 1571  
Qy : : : : :  
7166 GGTGTCAGGAGGACAGGAGCAGATGCGCAGCAGCTGCGGAGGAGAGCAGCGGCTTC 7225  
Db : : : : :  
1571 lnAlaArgAlaArgGlyPheMetAlaArgLysArgPheGlnAlaLeuMetThrProGlu- 1590  
Qy : : : : :  
7226 AG-----CGGACGCTGAGGCGCGAGCGCGAGCGGAGCTGAGATGAGCGCTGAGG 7276  
Db : : : : :  
1591 -----MetAspLeuIleArg----- 1596  
Qy : : : : :  
7277 CTGAGCGCTCAAGCTGCTGCGCGAGATGAGCGCGAGCGCGCGCGCTGAGGAGG 7336  
Db : : : : :  
1597 -----GlnLysArgAlaAlaLysValIleGlnArgTyrTrpArgGlyTyrL 1612  
Qy : : : : :  
7337 ACGCCAGCGCTTCCGAGAGCGCGGAGGAGATCGGTGAGAGCTGCGCAGCGAGC 7396  
Db : : : : :  
1612 eu-----IleArgArgGlnLysH 1619  
Qy : : : : :  
7397 TCGCCACCAGGAGAGGTGACCTGCTGCGAGACTGCGAGATCCAGCGAGCAGAGTG 7456  
Db : : : : :  
1619 isGlnGlyLeuLeuAspIleArgLysArgIleAlaGlnLeuArgGlnGluAlaLysAlaV 1639  
Qy : : : : :  
7457 ACCATGATGCGCGAGCGCTGCGGAGCGCATCGCTGAGCTGCGAGCGTGAAGAGG- 7512  
Db : : : : :  
1639 alAsnSerValArgCysLysValGlnGluAlaValArgPheLeuArgGlyArgPheIleA 1659  
Qy : : : : :  
7513 -----AAGCTCCACAGGAGCGCGGCGGAGGAGCTGCTGAG-CTCA 7546  
Db : : : : :  
1659 laSerAspAlaLeuAlaValLeuSerGlnLeuAspArgLeuSerArgThrVal-ProHis 1678  
Qy : : : : :  
7547 AGTCTGAGGAGATGACAGCGTGCAGCAGCAGCAGCTGCTGCGAGGAGCGCGCGC- 7603  
Db : : : : :  
1679 LeuLeuMetTrpCysSerGluPheMetSer-----ThrPheCysTyrGlyIleMet 1695  
Qy : : : : :  
7604 -----TGACGAAAGCTTCTCTCTGAAAGGAGCAGCGCTGCTACAGCGGAGC 7651  
Db : : : : :  
1696 AlaGlnAlaIleArg 1700  
Qy : : : : :  
7652 GCTTCATCGAGCAGG 7666  
Db : : : : :  
RESULT 15  
US-09-917-800A-1505  
; Sequence 1505, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880



/ PRIOR FILING DATE: 2000-11-02  
/ PRIOR APPLICATION NUMBER: US 60/290,029  
/ PRIOR FILING DATE: 2001-05-11  
/ PRIOR APPLICATION NUMBER: US 60/290,645  
/ PRIOR FILING DATE: 2001-05-15  
/ PRIOR APPLICATION NUMBER: US 60/292,336  
/ PRIOR FILING DATE: 2001-05-22  
/ PRIOR APPLICATION NUMBER: US 60/295,798  
/ PRIOR FILING DATE: 2001-06-06  
/ PRIOR APPLICATION NUMBER: US 60/297,457  
/ PRIOR FILING DATE: 2001-06-13  
/ PRIOR APPLICATION NUMBER: US 60/298,884  
/ PRIOR FILING DATE: 2001-06-19  
/ PRIOR APPLICATION NUMBER: US 60/303,459  
/ PRIOR FILING DATE: 2001-07-09  
/ NUMBER OF SEQ ID NOS: 1740  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 1505  
/ LENGTH: 15231  
/ TYPE: DNA  
/ ORGANISM: Rattus norvegicus  
/ FEATURE:  
/ OTHER INFORMATION: Genbank Accession No. US20020119462A1 X59601  
/ US-09-917-800A-1505

Alignment Scores:  
Pred. No.: 2,3e-19 Length: 15231  
Score: 317,50 Matches: 367  
Percent Similarity: 35.11% Conservative: 277  
Best Local Similarity: 20.01% Mismatches: 670  
Query Match: 3.34% Indels: 522  
DB: 81 Gaps: 81

US-09-914-698-1 (1-1861) x US-09-917-800A-1505 (1-15231)

QY 52 LysThrLeuGlnLeuLysSerProThrGlyAlaGlyLysThrMethySerValSer 71  
Db 3565 AAAGTTCTGCTTGCAGAGCGCTCAGCTGTCACCACTGCGCTCGAGTTGGAA 3624  
QY 72 AlaAlaValGlnGlnLysLys-ArgMetSerAlaAlaAlaAlaProSerLysGlnTh 91  
Db 3625 TTGACCTGGCGAGCTGGAGAACAGG---TCAGAGCGCTGTCGCA-----TCTAC 3672  
QY 91 rTPrArgValThAlaProSerArgProAlaAlaTAlaAlaHisProProGlnAlaPr 111  
Db 3673 TTGGAGA-----AACTCAAGACCATCAGCTTGTAATTGCGAGTACCAGGGGGCT 3723  
QY 111 oLeuValGluLysAsnValTyrLysThrProGlnGluProValTyrLysSerProG1 131  
Db 3724 -----GAGGAGGTGCTTAAACACACAGGAGGACACCTGAAGGAGGCCCGGCT 3773  
QY 131 nProArgSerLeuLysGluAsnLeuSerProMetThrProGlyAsnLeuAspVal-- 150  
Db 3774 GCCTGCCACACTCCAAG-----CTCGAAGTCAAC 3803  
QY 151 -----IleAspAsnLeuArgPheThrProLeuThrGluThrArgGlyLysGlyG1 167  
Db 3804 CAAGGCTTCACTAAGAGAGTGGCGG-----GCCAGCGGAGGCACA 3845  
QY 167 nAlaThrIlePheProAspAsnLeuAlaAlaTPrProThrProThrLysGlyAsnVa 187  
Db 3846 GCAGCCTGTATTC----- 3858  
QY 187 LysSerCysAlaAsnAspMetArgProArgArgIleThrProAspAspLeuGluAspG1 207  
Db 3859 -----AACACCTACCA-----GATCAGCTGAGGGGGC 3887  
QY 207 nProAlaThrAsnLysThrPheAspValLysHisSerGluThrIleAsnLysSerLeuAs 227  
Db 3888 ACAGGAGTTGGTAACGGCTACAGCAGCGCGCATGGTGAGCGG----- 3930  
QY 227 pThrLeuAspCysSerArgIleAspGlyGlnProHisThrProLeuAsnLysThrThr 247  
Db 3930 ----- 247

Db 3931 -GACGTGGAAGTAGAGCGCTGGCGAGAACGTGTCTCACTGCTGTGGAGCCCTGGCAGGC 3989  
QY 247 rIleValHisAlaThrHisThrArgAlaLeuAlaCysIleHisGluGluGlyProSe 267  
Db 3990 TGTCTAGCCAGACACTGATGTCGG----- 4014  
QY 267 rProProArgThrProThrLysSerAlaIleHisAspLeuLysArgAspIleLysLeuVa 287  
Db 4015 -----CAGCGGAGCTTGAACAGCT 4034  
QY 287 lGlySerProLeuArgLysTyrSerGluSerMetLysAspLeuSerLeuLeuSerProG1 307  
Db 4035 GGGCGGCCAACTTGGTACTACCTGAAAGTGGCGATCCGCTGAGCTCCTGGCTGCAGGA 4094  
QY 307 nThrLysTyrAlaIleGlnGlySerMetProAsnLeuAsnGluMetLysIleArgSerI1 327  
Db 4095 TGCCAAAG----- 4101  
QY 327 eGluGlnAsnArgTyrTyrGlnGlnGlnGlnIleGlnIleLysAlaLysAspLeuAsnSe 347  
Db 4102 -----AGCCGGCAAGAACAGATCCAGCTGTGCCAATAGCCAA 4139  
QY 347 rSerSerSerSerGluAlaSerLeuAlaGlyGlnGln-----GluPheLeuPheAs 364  
Db 4140 CAGTCAGCTGCACGAGAACAGCTGCCAGGAGAGGCCCTGCTGGAGGAGATTGAGCG 4199  
QY 364 nHisSerGluIleLeuAlaGlnSerSerArgPheAsnLeuHisGluValGlyArgLysSe 384  
Db 4200 CCATGTTGAGAGAGTTGAGGAGTGCCAGAAAGTTTGTAAAGCAGTACATC-----AATGC 4253  
QY 384 rValLys-----GlySe 388  
Db 4254 AATCAGGACTATGAGTCCAGCTGATCATCTCAAGGCTCAGCTGAAOCTGTGGCTC 4313  
QY 388 rProValLysAsnProHisLysArgArgSerHisGluLeuSerPheSerAspAlaProSe 408  
Db 4314 CCGCGCAAGAGCCCAAGTTTCACTGATCGGAGAGGCTCATCCAGAGTACGTGGA 4373  
QY 408 rAsnGluSerLeuTyrArgAsnGluThrValAlaIleSerPro----- 442  
Db 4374 TCTGCTGTACAGCTACAGTGAGTGACCACTCAGCAGTCAGTACATCAAGTTTCATCAG 4433  
QY 423 -----ProLysLysGlnArgValGluAs 430  
Db 4434 TGAGACACTGCGCGCGATGGAAGAGAGAGCGGCTGCTGAGCAACAGCGCGCAGGGA 4493  
QY 430 p---ThrThrLeuProArgSerAlaAlaProAlaAsnAlaSerAlaArgSerSerAl 449  
Db 4494 GCGGAGCGCTGCGCGAGTGGAGCGCGCTGGAGAGCAGCGGCGGAGTGGCTGAGGC 4553  
QY 449 aHisAlaTPrProHisAlaGlnSerLysLysPheLysLeuAlaGlnThrMetSerLeuMe 469  
Db 4554 CCATGCCAGCGCCAGAGCAGCGCC---GAGCTGGAGGACGAGAACTGCGCGCGCAT 4610  
QY 469 tLysLysProAlaThrProArgLysValArgAspThrSerIleGlnProSerValLysLe 489  
Db 4611 GCAGGAGAGGTGACCGCGCGGAGGAGCGCGGTGGACGACAGCAACAGAAAGCGCAG 4670  
QY 489 uTyrAspSerGluLeu---TyrMetGlnThrCysIleAsnProAspProPheAlaAlaTh 508  
Db 4671 CATCCAAGAGGAGCTGACGATCTGGCGCAAGCTCAGAGGCGAGAGATCCAGCCCAAGGC 4730  
QY 508 rThrThrIleAspProPheLeuAlaSerThrMetTyrLeuAspGlu----- 523  
Db 4731 CCAGCAGGTGGAGCTGACAGCGCGCATCGCATTTAGGAAAGAGATCCGCGTAGT 4790  
QY 524 -----GlnAlaValAspArgHisGlnAlaAspPheLysLysTrpLeuAsnAl 539  
Db 4791 CCGTCTGAGTAGAGCAACACTGAGCGTGGCGGTGGAGGGCGGAGATGAGTCAGGC 4850  
QY 539 aLeuValSerIleProAlaAspLeuAspAlaAspLeuAsnAsn---LysIleAspValG1 558  
Db 4851 TCTGCTGACCGGCTGAGGAGGAGCAGAACACAGAGCGGCGAGGCTCAGGAGGAGCCGA 4910





*This Page Blank (uspto)*